1-1-1989

Human factors implications of conceptual design representation in very large databases.

Edward Harding
University of Massachusetts Amherst

HUMAN FACTORS IMPLICATIONS
OF
CONCEPTUAL DESIGN REPRESENTATION
IN
VERY LARGE DATABASES

A Dissertation Presented

by

EDWARD HARDING

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1989

School of Management

HUMAN FACTORS IMPLICATIONS
OF
CONCEPTUAL DESIGN REPRESENTATION
IN
VERY LARGE DATABASES

A Dissertation Presented

by

EDWARD HARDING

Approved as to style and content by:

Van Court Hare, Jr., Chairperson of Committee

W. Bruce Croft, Member

Robert A. Nakosteen, Member

James E. Owers, Member

Ronald J. Karren
Ph.D. Program Director
School of Management

ABSTRACT

HUMAN FACTORS IMPLICATIONS

OF

CONCEPTUAL DESIGN REPRESENTATION

IN

VERY LARGE DATABASES

September 1989

EDWARD HARDING, A.B., MIDDLEBURY COLLEGE

M.B.A., DARTMOUTH COLLEGE

Ph.D., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Van Court Hare, Jr.

The continuing trend of greater availability of very large databases (VLDBs) provide both opportunities and challenges to users. One user challenge is the acquisition of proficiency in navigating through the VLDB in a minimum amount of time, with the least amount of effort, while still being able to locate the information of interest.

This work regards the interface between users and large databases. Empirical data are derived from a human factors experiment in which unsophisticated users execute queries on a database, thereby generating measures of performance. The analysis of results contradicts assumptions about one form of

representation being more effective than another form in communicating the structure of the data to users.

The thesis is that both the graph and the table forms are equally effective as documentation, and that differences in query performance are more likely a function of individual differences. It is suggested that the conceptual design representations currently used to describe real world databases are selected, not by real measures of performance, but by other factors, particularly personal taste, and probably others, such as institutional requirements, or systems constraints.

The conclusion suggests that multiple conceptual design representations be available for multiple users.

TABLE OF CONTENTS

ABSTRACT iv

LIST OF FIGURES viii

Chapter

1. INTRODUCTION 1

A. Introduction 1
B. Thesis 1
C. Motivation 2
D. The Users 2
E. The Database Market 3
F. Opportunities and Challenges 3
G. Management Perspective 4
H. Psychological Issues 5
I. Representations 7
J. Current Standards 8
K. Summary 12

2. REVIEW OF THE LITERATURE 14

A. Historical Perspective 16
B. Nature of Data Models 18
C. Function of the Model 18
D. Classification Schemes 19
E. Graphs 20
F. Records 20
G. Normalization 21
H. Structural Elements 22
I. Representation Research 30

3. THE HYPOTHESIS 37

A. Hypothesis Development 37
B. Hypotheses Statements 43
C. Significance of the Research 45

4. THE EXPERIMENT 48

A. Experimental Approach 48
B. The Design of the Experiment 53
C. The Database 58
D. The Queries 60

TABLE OF CONTENTS (continued)

5. DATA ANALYSIS AND CONCLUSIONS 70

A. Pre-Test 70
B. Demographics 71
C. Test Results - Efficiency 76
D. Test Results - Effectiveness 78
E. Threats to Internal Validity 80
F. Threats to External Validity 81
G. Conclusions 84

6. APPENDICES 88

A. Database Development Process 88
B. Model Attributes 90
C. Selected Models. 93
D. Test Instrument 108
E. Raw Data 117
F. The Database 127

BIBLIOGRAPHY 162

LIST OF FIGURES

1. Example of an SQL Query 10

2. Example of a QBE Query 11

3. Example of an E-R Diagram 12

4. Two Representations 15

5. Graphical Representation 20

6. Record Layout 21

7. Relationships 24

8. Attributes 25

9. Aggregation and Generalization . . 27

10. Existence 29

11. Sample Table Documentation 49

12. Sample Graph Documentation 51

13. Hierarchical Structure 94

14. Network Structure 95

15. Relational Structure 96

16. Cardinality in the E-R Model . . .100

CHAPTER 1

INTRODUCTION

A. Introduction

This work regards the interface between users and computerized databases. Empirical data are derived from a human factors experiment, described in Chapter Four, in which unsophisticated users execute queries on a database, thereby generating measures of performance. The analysis of results contradicts some myths and assumptions about the relative merits of the visual representation of two generally accepted models.

B. Thesis

Briefly, the myths suggest that certain visual models are inherently better than others. However, the thesis of this work is that two popular visual forms (the Graph and the Table) are equally effective in communicating the essence of data structure for the intermittent user.

Proponents of particular models (see Appendix C) may suggest otherwise, either implicitly or explicitly, yet the results of this experiment leads to the conclusion that any differences in performance must be due to individual differences.

C. Motivation

The motivation in pursuing this work was to add to the knowledge of what constitutes "quality" in database design, particularly from the viewpoint of the casual user. As a periodic user of a variety of databases myself, over the years I've become increasingly aware of different visual representations of data structure. Perhaps my impatience outweighs my attention span, so when looking for data in an unfamiliar database, I want the quick overview that will yield immediate results, and don't want to have to wade through pages of documentation.

D. The Users

Satisfying this researcher's impatience is an insignificant issue compared to the larger issue of addressing the needs of the millions of other intermittent users. They are, for example, the managers, analysts, and researchers who have occasional need to access very large databases.

The presumption is that these users access a variety of databases, but only infrequently, and that they need some form of documentation to comprehend the organization of the data in order to find items of interest, or to formulate queries.

E. The Database Market

While the number of intermittent users is difficult to measure, the scope of the database market is known. The worldwide market for database management systems (for mainframes and minicomputers) for 1987 was estimated to have been $3.1 billion with growth expected to reach $7 billion by 1992 [INGRES, 1988].

F. Opportunities and Challenges

This continuing trend of greater availability of very large databases (VLDBs) provide both opportunities and challenges to casual users. The opportunities include those activities which rely on (or are able to take advantage of) quick access to a high volume and wide range of data. Marketing research, demographic analysis, office automation and the examination of financial markets are typical examples of these activities.

Concurrent with these opportunities are a set of challenges for the users of VLDBs. One user challenge is the acquisition of proficiency in navigating through the VLDB in a minimum amount of time, with the least amount of effort, while still being able to locate the information of interest. Some further discussion as to the nature of this navigation activity is presented later in Chapter Three.

G. Management Perspective

The business's investment in a VLDB is based on an assumption that the data contained within the database will be utilized in a furtherance of the objectives of the management [Nolan, 1974]. As with other business assets, the value of having data is not solely inherent in the data itself, but in the potential of that resource to be utilized.

Using the Resource: Utilization of the data resource is typified by quick and effective access to large amounts of information through batch processing, on-line transaction processing, or one time inquiries. This work addresses problems associated with the last of these, one time inquiries, performed by part time users. It is worth mentioning that the one time inquiry is a significantly different activity than batch processing (e.g. shorter lead times) and that part time users have significantly different requirements than programmers and systems analysts (e.g. easy access).

User Problems: A fundamental problem is that there is often so much data in a VLDB that part time users often encounter difficulties in interacting effectively with the database. Other problems may include:

1) excessive time to search for particular data items,
2) inability to express searches in appropriate language,
3) avoidance altogether of initiating a search,
4) an inability to see larger patterns within the database because of preoccupation with individual data items.

Management Problems: The above problems relate to the interaction between the user and the VLDB, and these challenges are, by definition, also problems for the management of the firm. If these problems can be addressed by way of making the user more comfortable with the database, then the users are likely to be more productive in meeting management's objectives [Thomas, 1987].

H. Psychological Issues

Influencing Factors: A variety of psychological factors have been suggested as possibly influencing a user's degree of compatibility with a VLDB. Here the concept of "compatibility" could be measured in terms of the ability to retrieve information quickly and accurately. For discussion purposes the psychological factors are gathered into the following groups.

1) Individual differences: (e.g. cognitive style, prior experience, attitude and ability [Ramaprasad, 1987]).
2) The physical environment within which the search process takes place: (e.g. level of distractions, discomfort, noise, and the ergonomics of the human/machine interface [Shneiderman, 1987]).
3) The display of the data: Display enhancement through the use of color, or hard copy availability (versus CRT display), may have an impact on the effectiveness and efficiency of the database retrieval process [Benbasat and Dexter, 1985].

The Representation Factor: This last issue, the display, or visual representation of the data, is investigated in this research. More specifically, the research seeks to determine if the choice of a particular visual representation of the data structure has an impact on the ability of the user to effectively conceptualize the logical structure of the database.

Cognitive Psychology: The underlying assumption that conceptual structuring exists within the human mind and that this structuring might have an effect on database usage, lies in the realm of cognitive psychology. A brief summary of these issues follows.

Conceptual structures are a form of cognitive simplification. Simplification refers to the process of the human mind perceiving external patterns, and matching the perceptions to a set of similar, previously recognized, structures in the mind. Through this process of associating perceptions to previous recognized patterns, the human mind is able to give meaning to an image [Sowa 1984].

Assuming the credibility of the previous explanation of the mental construction of meaning, it is plausible to make a further logical leap. That is, for a database to have meaning to a user, it is necessary for the user to be able to conceptualize the logical organization of the entities within the database.

I. Representations

Popular Representations: There are a variety of different representations of conceptual structures for data. For example, flowcharts, spreadsheets, card files, and menus might all be considered data representations. Kent suggests that tables and graphs are two significantly different representations of data structures [Kent, 1978]. The use of tables, as a standard for data structure representation, has been firmly established by the proliferation of the Relational data model [Codd, 1970].

Alternately, the use of graphs is developed in the discipline of graph theory, a branch of abstract algebra, which investigates the notion of entities (or nodes) connected by arcs (edges or lines) [Harary and Norman, 1953]. The notion of using graphs to represent data sets has taken several forms, with one of the most popular being the Entity-Relationship (E-R) model [Chen, 1976].

This research uses the structures based upon
1) the table concept typical of those used in popular query languages developed concurrently with the Relational data model and,
2) the graphical format based partially on classical graph theory and similar to the graphical representation used in the E-R model.

J. Current Standards

Of the worldwide market for database management systems (DBMSs) in 1987, 61 percent of the total $3.1 billion market was attributed to products based on the Relational model [INGRES, 1988]. Further, by 1992, this model's predominance is expected to rise to 85 percent of the anticipated $7 billion total market.

Further, current commercial databases have evolved to the point of having created two de facto standards for the expression of queries made against them. By name, the two standards are the Structured Query Language (SQL) and Query By Example (QBE), and by nature these two standards are significantly different, especially in light of their relevance to this research.

SQL: Derived from SEQUEL, the structured English-like query language, SQL, [Chamberlin and Boyce, 1974], has attained the most commercial success as a high level format for expressing queries [INGRES, 1988]. The success of the language may be attributed to its flexibility in being able to use it with Cobol, PL/1, Fortran, and assembler application programs or to use it alone in an interactive mode [IBM, 1987].

Nonetheless, SQL's importance is discounted in this research for the following reason. It may well be the popular format for expressing a query, but SQL is not significant as a format for the representation of the underlying data structure of the queried data. Indeed, the language is more procedural in nature (see Figure 1), and does not include the facility of providing a visual representation of the data structure.

```
select PART-NO, DESCRIPTION
  from PART
  where PART-NO=
  (select C-PART-NO from ORDERS
    where C-INVOICE-NO=1234)
```

Figure 1 - Example of an SQL Query

QBE: In contrast to SQL, Query-By-Example (QBE) as defined by Zloof of IBM [Zloof, 1975] provides not only a high level facility for the expression of queries, but it also provides a visual representation of the data structure of the database. Currently QBE is offered as an interface to such commercial products such as PARADOX, DB2, dBase IV, and ORACLE.

The QBE format is characterized by a row and column format that is essentially the tabular structure (being addressed by this research) and is reminiscent of the classic record and field layouts used in databases. Queries are activated by filling in skeleton tables (see Figure 2) displayed on a CRT with examples of a query or update operation which are then associated with actual tables of the database [Zloof, 1982]. A distinction is made between a "constant element", and an "example element" (variable), where "example elements" are underlined and "constant elements" are not. These example elements are used to define links among two or more rows of the same table or among different tables, so, by definition, example elements are superfluous in queries where links are not employed [Zloof, 1985].

SUPPLIER	PART_NAME	PART_NO
Acme	Button	1234

Figure 2 - Example of a QBE Query

Testing QBE: Some psychological experiments have been performed on QBE [Thomas and Gould, 1975]. Results indicated that in a few hours of training, non-programmers were able to pose relatively complex queries. The test used total training time and number of correct queries as the measure of success of the language.

Graphical Modeling: Yet for expressing perceptions of reality in a more structured form, another model, Chen's E-R model (see Figure 3) has emerged as the dominant representation. The model has gained credibility in both academic research and in commercial automated tools for systems development because of its versatility and clarity in expressing complex relationships observed in the real world.

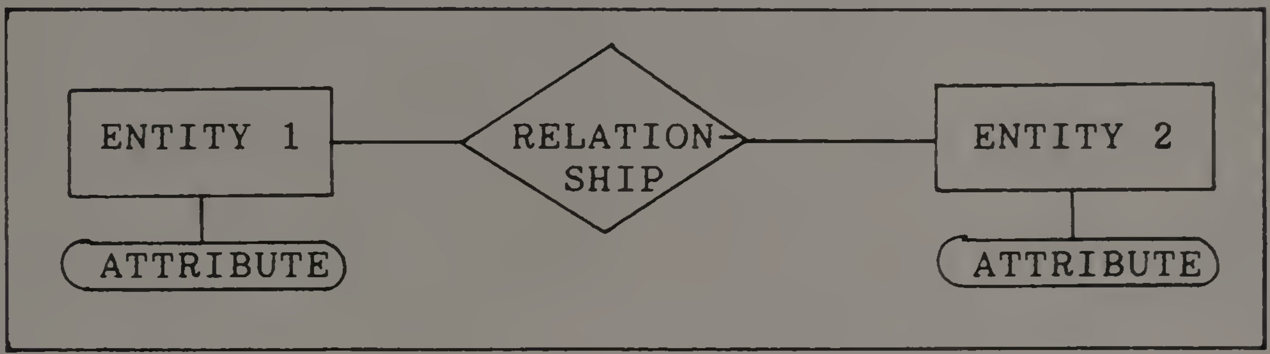


Figure 3 - Example of an E-R Diagram

Use of Graphs: The E-R model is most commonly used to express reality prior to building the database (the systems development cycle is more thoroughly discussed further on). Recently, however, E-R modeling tools have been developed to reverse engineer systems [Chen, 1988]. That is, they are used to take existing schemas of databases and convert them to E-R diagrams. This suggests they have value is a form of documentation of existing databases, a viewpoint adopted in this research.

K. Summary

It is within this context of commercial firms, very large databases, part time users, and different models for the representation of data organization that this research evolved. Curiosity about the relative merits of the different proposed data models added to the original motivation for pursuing this experiment. The remainder of the paper is structured as follows. Chapter Two discusses the nature of conceptual data models and the querying process of this work. It also notes some of the significant contributions to the field of database design and experimentation that brought the field to its current state.

In Chapter Three, the hypothesis statements of the thesis are more fully specified. Chapter Four describes the experiment that generated the data in support of the hypotheses and Chapter Five provides the analysis of that empirical data and uses the results to draw some generalizations from the data and links them to some overall conclusions. Chapter Six is an appendix of reference material, including the actual database used in the experiment, a facsimile of the test instrument, and the raw data of the tests. Finally, Chapter Seven contains the bibliography.

CHAPTER 2

REVIEW OF THE LITERATURE

The notion of testing different representations of conceptual data structure, as addressed in this work, is built on a foundation of previous work in data models, data representation, and database experiments.

The distinction between data models and data representation deserves some emphasis. A data model defines the rules by which data elements must be structured and manipulated [Tsichritzis and Lochovsky, 1982]. On the other hand, data representations are the visual conventions used to express that data structure.

A simple illustrative example of the same structural data model having different data representations is described here. We assume that the example data model defines three similar entities, each of which has a relationship with three other entities. A network data representation of this structural model appears on the left in Figure 4, while a matrix data representation of the same structure appears on the right. Although many other representational conventions also exist, these two illustrate the point.

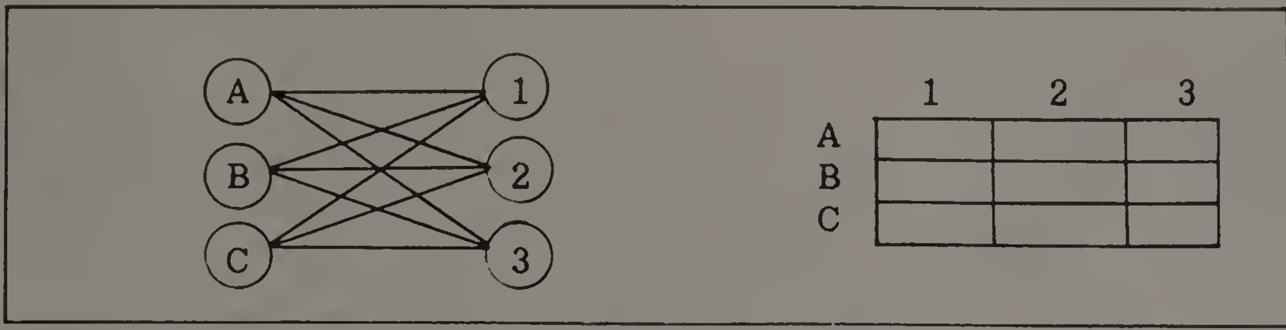


Figure 4 - Two Representations

The broader issue of which data models best represent the realities of the enterprise could not be fully addressed without first some resolution as to how the structures are best visually represented. The experiment of this research deals specifically with the second issue, differences in representation of the structures, not differences in structure, but is built on the foundation of previous work in the area of conceptual data models.

Much of the conceptual data modeling literature focuses on the structural issues with less emphasis on the representation of the structures. Discussion of the structural contributions may be found in Appendix B, and C.

Consequently, this chapter includes a review of the historically significant developments in data modeling, the concurrent representational issues, and previous experimental work which influenced the general experimental design of this research.

A. Historical Perspective

The introduction of computers into organizations in the 1960s prompted the first attempts to model the enterprise in terms of machine maintained data. The process involved users defining the requirements to programmers who wrote the code. Generic DBMSs were subsequently developed during this period and were typically of the hierarchical or network format. Representation of the hierarchical model drew heavily upon graphical trees, while the network models were illustrated by Bachman diagrams.

In 1970, Codd published his work on the Relational model based on relational set theory [Codd, 1970] which provided the basis for subsequent implementations of that model. The construct of a "data model" is attributed to this work in which Codd proposed a method for representing the real world via object definitions, operations on the objects, and rules to enforce the integrity of the data [Date, 1985].

The Relational model addressed a perceived need to develop a more theoretically pure approach to the organization of data derived from a complex world which often lacks obvious structure. Data models thus fulfilled the need for a critical link in the process of expressing the enterprise as a database.

Meanwhile, the Conference on Data Systems Languages (CODASYL) Data Base Task Group (DBTG) formalized standards for the already implemented network format [CODASYL, 1971]. Debate arose as to the relative merits of these first generation models (the hierarchical, network, and relational) [Codd and Date, 1974], but the issue was diffused somewhat in the middle 1970s with the introduction of several conceptual data models. In retrospect, Chen's Entity-Relationship model [Chen, 1976] has provided perhaps the greatest influence as a standard approach to conceptual data modeling, perhaps because of its elegant simplicity.

With the different focus of these "second generation" models, the debate was allowed to splinter into two sets of criterion for evaluation. The first set of issues, machine performance and implementation feasibility, were left to the first generation models. The second set of issues, capturing the true semantic meaning of the data of the real world, and doing this with natural, or intuitive, representation, was assumed by the newer conceptual data models [Lum, 1978].

B. Nature of Data Models

The classical approach to database development requires that the realities of the enterprise, as perceived by end users, be communicated to systems analysts. It is the analyst who maps reality to a structured conceptual data model. Appendix A describes this process in greater detail.

Quality models need to be both simple enough to be comprehensible as well as expressive enough to communicate complex relationships of reality. Appendix B provides more depth on these model attributes and Appendix C identifies and describes selected models.

C. Function of the Model

The traditional function of the conceptual data model is to provide a stepping stone between the often amorphous elements of reality, and the stringent structural requirements of most DBMSs during the development of the database. A secondary function of the conceptual data model is to provide documentation to casual users of existing databases, to provide an illustration of the database structure, so that queries may be performed.

D. Classification Schemes

With the diversity of data models, it is helpful to establish a framework, or taxonomy, in order to compare and contrast them by nature [Kerschberg et al., 1976].

Tsichritzis and Lochovsky note the confusion which is inherent in the attempt to compare and contrast the relative merits of the various data models. They contend that the confusion is due, in part, to the fact that data models are defined rather loosely, thus the concepts cannot be completely understood in a single framework.

They do draw a distinction between "strictly typed" data models in which each datum must belong to a category, and "loosely typed" in which individual pieces of data may exist independently [Tsichritzis and Lochovsky, 1982]. But beyond that, because the distinctive features of the various models are often conceptually at right angles to one another, these models often escape a neat comparison.

Nonetheless, frameworks have been proposed [Chen, 1981], [Date, 1985], which make a clear distinction between the n-ary and binary relationships, as more fully discussed in Appendix C.

Aside from the binary versus n-ary dichotomy, another split in approaches to modeling exists between record-like and graph-like models [Kent, 1978].

E. Graphs

In the graph approach, there exists only one occurrence of a given object and all its relationships radiate from that single occurrence (Figure 5). The pure mathematical study of graphs, as in graph theory [Harary and Norman, 1953] establish many of the representational primitives adopted in conceptual data modeling, notably in the expression of the E-R model.

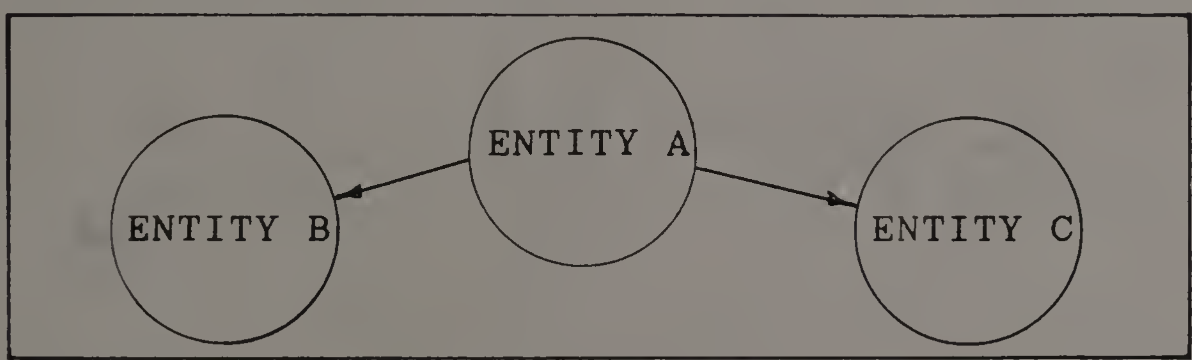


Figure 5 - Graphical Representation

F. Records

The record-like school involves replicated references and is more closely associated with table representation of data as manifested in the Relational model. The record (or table) oriented approach to data modeling has been criticized as fostering an asymmetry in the treatment of relationships [Kent, 1978]. That is, in record format, one-to-many relationships are given fundamentally different treatment from many-to-many relationships, the relationships being defined through the process of normalization (Figure 6).

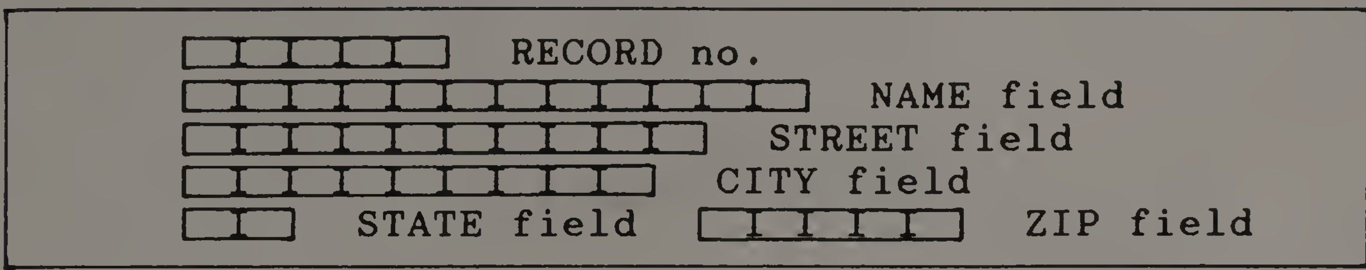


Figure 6 - Record Layout

G. Normalization

While normalization has the advantages of maintainability by virtually eliminating the chances of redundant (or conflicting) data, it does lack the expressiveness of the graphical approach by masking the relationships among entities.

This deficiency, or "lack of semantic quality control" [McFadden and Hoffer, 1985] is particularly apparent when a user attempts to reconstruct logical relationships from existing tables. Since each table exists on its own, there is no guarantee that a cross-reference key will reference an existing record, since the Relational model has no construct or property to force this matching [McFadden and Hoffer, 1985]. This issue of "referential integrity" is currently being addressed in recent releases of relational DBMSs, but will remain a weakness of the generalized conceptual relational model.

H. Structural Elements

Reference has been made to the significance of not only the entities of the enterprise, but also of the interconnections, or relationships, between the entities. The nature of those entities and relationships constitutes the conceptual design, or structure of the data elements. Yet, the term "data structures" may have distinctly different emphasis depending on the context. For example, one context in which "data structure" is often used is at a lower (closer to the machine) level, and deals with where and how data elements are stored and retrieved.

However, for this research "data structures" refers to the relationship between the data elements as defined by the rules of a particular conceptual data model, and thus represent a "higher level" context, or one closer to the user and more reflective of real meaning of the data elements. The following discussion addresses the particularly significant structural attributes, or conceptual data model primitives.

The Entity: The entity, or object, is the atomic piece of data, the thing of interest. Entities may be tangible objects (people, places, things) or abstractions such as classes of objects or events. Langefors suggests that the four types of critical information that may be recorded about an entity include the name of the entity, the property of the entity, the value of the property, and a time reference at which the information was valid [Langefors, 1977]. All conceptual models have a construct for at least the entity, and it is usually the element of most attention.

The representation of the entity in the E-R model is a labeled rectangle, or in more generalized graphics, simply a circle or node. In a table format, each row is an entity, often with the first column of the row containing a unique identification for that row.

The Relationship: The relationships (also known as the associations) among multiple entities, or among classes of entities, provide the basis for the structure, or schema, of the data model. The schema is a structured representation for a particular data model application which identifies the classes, or categories by name, the properties of the categories, and the relationships between the categories [Hammer and McLeod, 1981].

Relationships may take the form of hierarchies, such as aggregation and generalization hierarchies. They may also include notions of existence dependency, in which the existence of one entity is dependent upon the existence of another entity. These specialized forms of relationships are discussed below in more detail.

In the graphical format, relationships are shown as lines, or arcs, drawn between entities, and in the classical E-R diagram, a diamond on the line identifies the relationship. In the table format relationships are established by the inclusion of key fields from other entities (Figure 7).

CUSTOMER

NO	NAME
036	Chen
376	Dexter
709	Kent

ORDER

NO	CUST_NO
001	404
002	036
003	709

Figure 7 - Relationships

The Attribute: The attribute (also known as characteristic or property) of the entity is comparable to the Langefors property mentioned above. Attributes may be associated with relationships as well as with entities, depending on the conceptual model in use.

Attributes may have unique values as a function of their particular entity, or may be functionally dependent upon the value of other attributes, either of the same entity, or upon attributes of other entities. The sorting out of the functional dependencies is partially the basis for Normal Form Theory, which plays a critical role in the Relational data model.

Attributes are represented as lesser nodes (usually ovals) in the graph form, and in the table form, the columns contain the values of each attribute type (Figure 8).

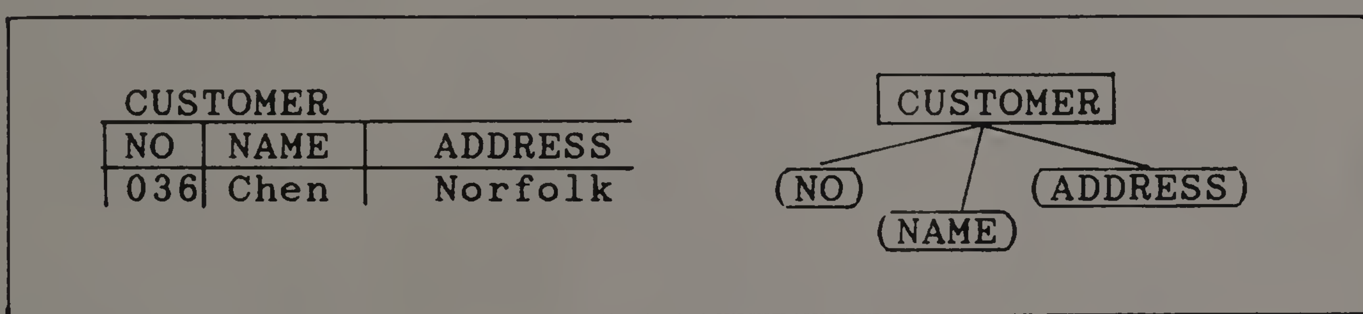


Figure 8 - Attributes

Aggregation: The abstraction of aggregation turns a relationship between objects into a higher level object, the aggregate object [Smith and Smith, 1977]. In some cases, the combined sum of attributes are said to be aggregated into the higher object.

These aggregate hierarchies are handled, for example, in Normal Form Theory where objects and their attributes are successively reduced. This reduction, or reverse aggregation, has been referred to as stepwise refinement, as in, "Gaul is divided into three parts".

Two types of aggregation have been proposed [Codd, 1979] which makes the distinction between a subtle variation found within instances of all aggregations. One type, Cartesian aggregation, includes the more straight-forward cases, such as the creation of an entity by references to the combination of all its attributes or, similarly, the aggregation of association and all its homogeneous participants [Date, 1983].

Cover aggregation, on the other hand, refers to the grouping together of heterogeneous entities into a higher level object, in accordance with some kind of membership criterion [Date, 1983]. Sometimes known as "set operator defined subclasses", the cover aggregation requires no natural member attribute predicates to be included in the sub-class [Hammer and McLeod, 1981].

Generalization: The process of generalization is the abstraction process which enables a class of individual objects to be thought of generically as a single name object [Smith and Smith, 1977]. As the higher "generic" object is a more generalized form of the "descendant" objects, attributes of the generic object may be inherited by the lower part objects. A mutually exclusive group of generic objects sharing a common parent is a "cluster", but it is not always true that descendant objects will be mutually exclusive.

Within the concept of class intersection [Hammer and McLeod, 1981] refers to entities that belong to two (or more) data classes. Also, the Extended Relational model, RM/T [Codd, 1979], offers the notion of "alternative" generalization where an entity may be a conditional subtype of a set of supertypes [Date, 1983].

When aggregation and generalization are used together in a model, they may be graphically represented on orthogonal planes (Figure 9).

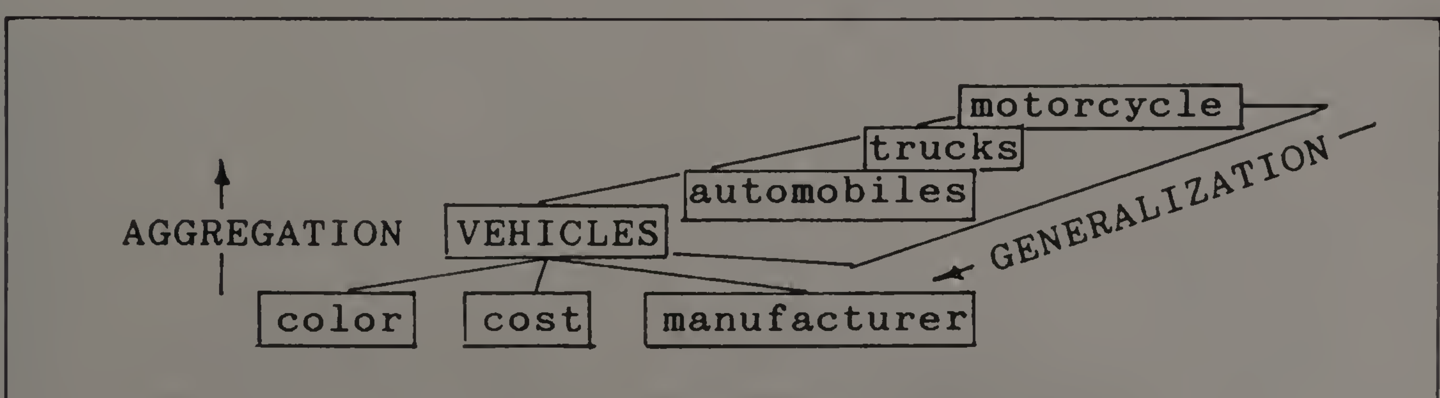


Figure 9 - Aggregation and Generalization

The aggregation of objects, such as parts, are aggregated from bottom to top in a hierarchy of pieces, from sub-assemblies to whole objects on a vertical plane.

The generalization plane is represented as a horizontal plane, and from front to back, it provides a structure for subtypes to be generalized into higher, more forward supertypes. In a pure table format, aggregation and generalization must be inferred from the normalized table.

Existence: In cases of entity relationships where the very existence of an entity is conditional upon the existence of another entity or entities, then an existence dependency is said to be present. An example would be the purchase of a customer being dependent on the existence of a record for that customer. Existence dependency is a type of referential integrity which may or may not be enforced by any particular DBMS.

A graphic form often chosen to specify the existence dependency is simple and unambiguous - a directed line (shown as a line with an arrow) from the dependent to the determining entity type (Figure 10).

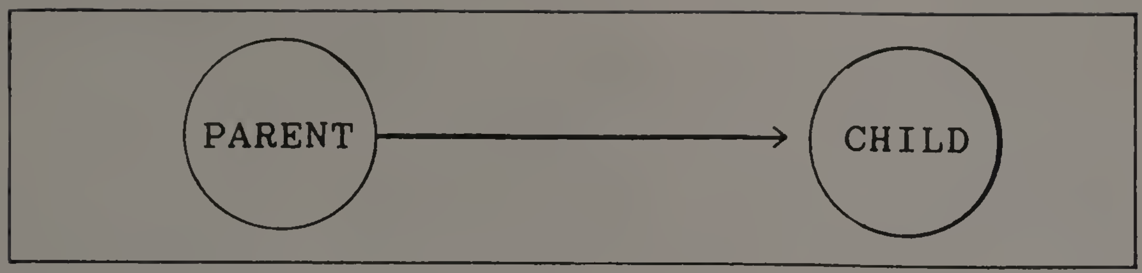


Figure 10 - Existence

A network of these existence dependencies may be arranged to form a "directed" graph [Harary and Norman, 1953]. However, for the table, there is currently no widely accepted visual convention for expressing existence, although the facility may be supported within the DBMS.

Polymorphism: Polymorphism is the facility which allows an entity to assume more than one form within a single database. The same construct has also be referred to as relativism [Hammer and McLeod, 1981], which emphasizes that the class to which an entity belongs is relative to the particular user's view that is being regarded.

This accommodation of multiple views of the same data enables user to develop new perspectives on the data, but it demands flexibility of the database and creates logically redundant data. In a database schema that allows logically redundancy, the values of some of the data is derived from the values of other data.

The impact is to reduce the amount of manipulation otherwise required to generate often used data. Further, a tight integration of the multiple views must be supported by the DBMS in order to maintain the data integrity, which may otherwise be compromised by the redundancy.

I. Representation Research

Some authors suggest that while the different representations for data structure are somewhat arbitrarily chosen, they still have rigorous rules for each formalism [Tsichritzis and Lochovsky, 1982]. The issue of the implications of structural representation on performance is implied by several other works. For example, Kent suggests that "unfortunately, the choice of representation too often does show through to impact the way users use the information" [Kent, 1978].

Representation formats that fail to clearly communicate semantic qualities can lead to user confusion and eventual frustration. As some research has demonstrated, users can become lost in a database, by virtue of not knowing where they are [Mantei, 1982].

Further, Nolan and Seward note that frustration may eventually lead to abandonment of the search or the entire system, and that this implies that an effective information system is one which requires relatively few expanded searches for information [Nolan, 1974].

The area of human factors experimentation as applied to information systems has focused more on hardware and displays, perhaps as an outgrowth of the early human factors work in airplane cockpit design [Ramsey and Atwood, 1979]. An example of the display experiments is illustrated in empirical research on human factors in data representation by Benbasat and Dexter.

Benbasat and Dexter: They established the influence of graphical and color-enhanced data representation on decision quality, decision making time, use of information, and user perceptions [Benbasat and Dexter, 1985].

One aspect of their research concentrated on comparative performance measures of output data represented in either a graph or table format. While superficially this appears similar to the work in hand, the nature of the information being expressed, and the meaning of graph and table formats were different than what is used here.

Specifically, the Benbasat and Dexter experiments regarded the output display of individual data elements (accounting information), whereas this work regards the display of the general structure of the data. Also, the Benbasat and Dexter graphics refer to Cartesian line charts, as in time series data, to illustrate the value and relationship of individual data elements.

For the work in hand, graphics refer to node and arc diagrams as suggested by mathematical graph theory and adopted to E-R modeling conventions. Similarly, a Benbasat and Dexter table refers to a formatted list of data values, whereas, in the context of this research, table refers to a generalized, record and field type layout.

Thus, the significance of the Benbasat and Dexter research, relative to this work, is not so much the specific results, but more in the approach to their hypothesis testing, which influenced the original experimental design of this research. Aside from their work, however, the amount of research done in the area of graphical versus non-graphical displays has been "disappointing" [Ramsey and Atwood, 1979].

Reisner: In the general area of queries on databases, Reisner notes six following areas of potential empirical work [Reisner, 1981].

1) Query writing tests, in which subjects translate English to a specific query language,

2) Query reading, in which subjects translate queries to English,

3) Query interpretation, in which users are given a query (in a query language) and a database and asked to find the data asked for,

4) Memorizing a database,

5) Complete problem solving with the database,

6) Question comprehension, in which subjects are given both English questions and the database and are asked to satisfy the query [Reisner, 1981]. It is within this last area of experiments that the research in hand is included.

One of the earliest attempts to quantify the intuitive nature (or ease of use) of a query methodology involved SQL [see also p.9] and a language similar in its functions, SQUARE. The nature of the conclusion was that programmers learned SQL "more completely" and non-programmers found SQL "easier" [Reisner, 1975].

Thomas and Gould: A related experiment was performed on QBE [see also p. 10] which concluded that QBE was easier to use than SQL based on measures of training time and accuracy of responses [Thomas and Gould, 1975]. The Thomas and Gould experiment was faulted for lack of scientific rigor by Greenblatt and Waxman who ran a more controlled test on the same issues [Greenblatt and Waxman, 1978].

Greenblatt and Waxman: While the Greenblatt and Waxman experimental design was not used as a model for the research in hand, there are some striking similarities in approach. The number of subjects for the Greenblatt and Waxman test was 25, and for this research, 30. Undergraduate students were used in both tests, and demographic data was noted, including gender, age, grade point average, and computer experience in both cases. The issue of the effect of intelligence on performance was considered an important variable to control in both the experiments, and the final measures included, among others, mean query times and percent correct.

Welty and Stemple: A similar experimental approach was taken by Welty and Stemple to determine whether the procedurality of a language affects its usability [Welty and Stemple, 1981]. Based on two languages, which differed mainly in procedurality, SQL and TABLET, they concluded that little difference existed for easy queries, but there were some differences to be noted on harder queries.

Lochovsky: When the notion of the data model is blended with the notion of query language, the experiments of Lochovsky must be noted. He compared the Relational, Hierarchical, and Network models, using their respective query languages, to conclude that the Relational model yielded higher performance measures [Lochovsky, 1978]. He qualifies his conclusions with the note that the data models and their accompanying data language were inseparable as a variable, such that much of the performance differences might be attributed to the language and not the model itself.

Brosey and Shneiderman: A decoupling of the model and the language factors was attempted by Brosey and Shneiderman who concluded that the Relational model appeared to be "convenient", whereas, in some cases the hierarchical model was "easier to use" [Brosey and Shneiderman, 1978].

Other Research: Giving more emphasis to the mental structuring of data elements, Durding, Becker, and Gould observed how subjects naturally structure data [Durding et al., 1977] and Broadbent and Broadbent noted the effects of individual differences (specifically educational background) on natural human data structuring [Broadbent and Broadbent, 1978].

Upon this theoretical background in conceptual data models, with references to the concurrent representational issues, and a sporadic but credible history of database experiments, the foundation of this research has been laid. It builds not on any specific research that has gone before, but borrows liberally from several, and covers one more of many heretofore unresolved issues.

As previous work in conceptual data models provides the structure by which information may be communicated to users, this work provides insight into how that structure may be visually presented.

CHAPTER 3

THE HYPOTHESIS

A. Hypothesis Development

The previous chapter noted that the function of the tables of the QBE format is to provide the database user with a visual representation of the structural relationships of the database elements. Also indicated was the point that the graphics of the E-R model are being used for the same purposes.

Some of the professional literature indicates a strong bias for either the graphics or the tables, although these biases are more implied than explicit. Research of integrity does not blatantly state that one form is better than another, because empirical data so far has been equivocal or inconclusive [Benbasat, Dexter, and Todd, 1986]. Nonetheless, the statement like the following are not unusual. "E-R Diagrams represent a case where the cliche 'A picture is worth a thousand words' actually holds true" [Flaaten et al., 1989].

Part of the motivation of the work in hand arose from a personal bias toward the graphical format and mild degree of professional offense taken as a result of what was perceived to be a purist attitude of the proponents of the table format. Therefore, this experiment was designed in hopes that the results would show that the graphical format provided superior documentation for users navigating through a database.

It is appropriate to diverge temporarily to more fully elaborate on the notion of navigation, as the concept plays such a crucial role in the hypothesis development and assumes a specific definition.

Navigation: The general meaning of navigating, "to steer a course through a medium" [Webster, 1975], of course, remains applicable. In the database environment, it refers to the journey from data element to data element, or from record to record, during the operations such as updates, searches, and joins. Most typically, the navigation is performed as an automated facility of a DBMS. For example, a DBMS structured after the CODASYL Network model uses a tightly structured logic to navigate (to select records) through the database [DG/DBMS].

A secondary, and less often used meaning of navigation refers to a user perusing the data, meandering amongst the labyrinth of information, or simply browsing the database. The latter term, browsing, is perhaps the most often used expression for this form of navigation.

An illustration of a manual form of this browsing navigation might be the flipping through a stack of 3x5 cards in search of particular data elements. Similarly, an illustration of an automated form might involve the use of a keyboard, mouse, touch screen, or other interactive device, to repeatedly scan records in a search, presumably in the attempt to satisfy either an informal query, or perhaps to retrieve specific data elements for a more structured query.

This secondary meaning of navigation through the VLDB has two variations. One variation is the active process of going from record to record, and the second is the more passive thought process of pondering the relationship between data elements, or "mentally navigating" through the database.

This research uses the term navigation in the browse sense. Further, it assumes that the measurement of the ability to actively navigate from record to record is a reasonable surrogate for the degree of ability to mentally navigate, or accurately conceptualize, the structure of the database.

Conceptualization: This conceptualization is not of the data elements themselves, but of the relationship between the data elements. These relationships are apparently challenging to express, and the ability to do so is a stringent test of quality of a representational form.

The idea that the representation form will influence navigational performance assumes the existence of a fundamental difference in the graphical versus tabular form of representation. While there may be several different dimensions for evaluation of the two forms, the critical dimension for this research is their relative ability to effectively communicate the various facilities proposed by the generally accepted conceptual data models. It is in the expression of these facilities that the understanding of the intended meaning of the data gets communicated to the user.

It is appropriate to emphasize the connection between the notion of representing data model facilities and the notion of doing searches on a manual database. This connection is central to the internal credibility of the experiment in this research, and the logical connection may not be immediately obvious.

While there is a strong connection between these two notions, it is not based on an assumption that the graph or table provides a direct aid to the user in navigating through the database as a sea chart would help a sailor. In fact, the navigation activity itself, in this experiment, is actually prompted instead by the listing of addresses of subsequent records noted on each card of the database. Subsequent records are determined by various different sorting criteria, as a database is indexed on different fields.

These direct navigational aids were available, and identical, on all sets of the experimental database, and thus both groups of users had identical direct navigational aids at their disposal. Without a doubt, some of the subjects of the experiment were able to gain proficiency in navigation more quickly than others, and not because of the representational form (the critical factor of this experiment), but because of their individual cognitive abilities.

The problem of individual differences in navigational proficiency is addressed in two ways. First, the randomly selected groups answered user profile questions designed to relate observed navigational proficiency to self assessed cognitive style. The results of these questions were then correlated with navigational performance.

Secondly, the design of the experiment attempted to isolate the navigational proficiency variable from the representational variable. This was achieved by having each group perform the query set twice, using one of the two respective representational forms each time.

So, whereas the "navigational skills" may not be solely determined by the documentation, the overall performance of query resolution is still believed to have been significantly influenced by the documentation. The critical difference between the direct and indirect influence is in the users ability to comprehend the relationships between the data elements as communicated by one of the two variant representational forms.

It is was originally thought that this comprehension would be aided (or impeded) by the particular representational form which was effective (or ineffective) in communicating the relationships. Further, it was assumed that this understanding would manifest itself in the user's ability to to decide what to look for within the database, and where it was likely to be, or similarly, how he or she was likely to be able to get there.

In conclusion, the research assumed that the query (a constant for the two groups) plus the representational form (one of two) would lead to a mental formulation of a search objective and strategy. And this formulation, combined with navigational skill would produce measures of search effectiveness and search efficiency.

B. Hypotheses Statements

In investigating the performance impact of the representational forms of the table and the graph, the researcher might have adopted one of several academic positions. For example:

1) that the Table should yield better performance results, consistent with its current wide acceptance and its straightforward simplicity.

2) that the graph should yield better performance results because of its richness in expressiveness and visual appeal.

3) that there will be no significant difference between the performance results of the two forms,

because they both communicate the same essential information.

It is the last position, with the benefit of hindsight, that is taken as my thesis. More specifically, while there are considerable differences between the two models (most notably in visual appearance and ability to express relationships), we suggest that the individual differences among the users will have a substantially greater and overriding impact on user performance than any differences between the Graph and Table formats.

The hypothesis statements that follow as a result of this position take the form of "there is no difference" and "there is a difference". More formally, they are:

The Null Hypothesis: There will be no statistically significant difference between the performance parameters of the graphic and tabular forms of structural representation.

The Alternative Hypothesis: The performance parameters of one form of structural representation will be statistically greater than those of the other.

The decision rules based on the hypothesis statements are as follows. If the null hypothesis is not rejected, then the results constitute supporting evidence to the position that graphical documentation is a valid form of user documentation (heretofore unsubstantiated in the literature or in practice). But it further suggests that claims of better performance with graphs continues to be without basis.

C. Significance of the Research

The continued growth in the development and use of very large data bases is indicative of their permanence in society. For batch processing applications and on-line transaction processing the structure of data is critical at the physical level (to address system performance issues), but the conceptual structure is less important, as long as it is eventually decipherable by the systems analyst.

This research does not address the design at the physical level of the database for two reasons. First, the physical structure of data is invisible to high level casual users, and is therefore unrelated to those users' ability to conceptualize the data. Second, currently available commercial databases are able to provide the logical links of entities that are necessary to replicate most structural forms proposed by the various models.

For other applications, however, there may be occasions to access the database on an ad hoc basis for analysis or research. In these instances, especially where there are...

1) massive amounts of data with complex relationships,
2) requirements for frequent ad hoc inquiries, and
3) users demanding quick response times,

the conceptual structuring and the representation of that structuring, may be critical in determining the effectiveness and efficiency of the database.

Effectiveness in this context is some measure of how well a human/system combination is able to find data elements, presumably through the execution of a query. Included in the concept of effectiveness are the concepts "recall" and "precision", where recall is the percent of the total relevant data items actually retrieved (as a result of the query), and precision is the percent of all the retrieved data items that are relevant. Relevance is a more elusive construct, not always easily recognized by machine (although quite consistently judged by humans). For this research, the notion of recall was used as the effectiveness measure.

Efficiency, in this context, deals with how quickly (assuming absolute confidence of levels of recall and precision) retrievals can be accomplished in light of some measure of storage requirements.

The issues of both effectiveness and efficiency are central to the significance of this research. Based on the results of this experiment, it is possible to add knowledge to what currently exists regarding the impact of conceptual structure representation on the effectiveness and efficiency of human interaction with VLDBs. By definition, these results further define what constitutes quality in database design and representation.

Even the results of no difference in performance is significant in defusing the argument of which form is inherently better. Further, that result might imply that several representational forms should be available to casual users.

CHAPTER 4

THE EXPERIMENT

A. The Experimental Approach

Aside from the theoretical construction of evidence to support the thesis of this research, an experiment was designed to measure performance differences between groups performing queries on a database with two variations of data structure representation.

Table Documentation: The first representation is in a normalized tabular format, typical of QBE tables (Figure 11). In this format, sample skeletal tables are illustrated as a form of documentation of data structure for users. On the tables, the table names (corresponding to record types) are noted along with the attribute names (i.e. fields, columns, domains) of the various tuples (ie. records or rows), and notation as to the significance of those attributes (for example, whether they are key fields or not).

While the table format has the advantage of being a closer representation of the form of the physical data record, the logical connections, or relations, are implied (by matching attributes), but not specifically noted.

```
CUST
CUST_NO (108) ¦ CUST_NAME (11) ¦ ADDRESS ¦
==============¦----------------¦---------¦
NEXT CUST_NO REC_NO ¦ NEXT CUST_NAME REC_NO ¦
- - - - - - - - - - ¦- - - - - - - - - - - -¦


EMPLOYEE
EMPLO_NO (93) ¦ SUPPL1_NO ¦ SUPPL2_NO ¦ SUPPL3_NO ¦ SUPPL4_NO ¦
==============¦-----------¦-----------¦-----------¦-----------¦
SUPPL1 REC_NO ¦ SUPPL2 REC_NO ¦ SUPPL3 REC_NO ¦ SUPPL4 REC_NO ¦
- - - - - - - ¦ - - - - - - - ¦ - - - - - - - ¦ - - - - - - - ¦
NEXT EMPLO_NO ¦
- - - - - - - ¦


ORDER
ORDER_NO (135) ¦ CUST_NO ¦ DATE (80) ¦
===============¦---------¦-----------¦
NEXT ORDER_NO REC_NO ¦ THIS CUST_NO REC_NO ¦ NEXT DATE REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - ¦ - - - - - - - - -¦


LINE
ORDER_NO ¦ LINE_NO (13)  ¦ QTY ¦ PART_NO ¦
=========¦===============¦-----¦---------¦
NEXT ORDER_NO REC_NO ¦ NEXT LINE_NO REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - ¦


SUPPLIER
SUPPL_NO (49) ¦ SUPPL_NAME (62) ¦
==============¦-----------------¦
NEXT SUPPL_NO REC_NO ¦ NEXT SUPPL_NAME REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - - -¦


PARTS
PART_NO (50) ¦ PART_NAME (8) ¦ EMPLO_NO ¦ SUPPL1_NO ¦ SUPPL2_NO ¦
=============¦---------------¦----------¦-----------¦-----------¦
NEXT PART_NO REC_NO ¦ NEXT PART_NAME REC_NO ¦
- - - - - - - - - - ¦ - - - - - - - - - - - ¦
```

Figure 11 - Sample Table Documentation

The second representation of data is in the graphical format similar to that proposed by the Entity-Relationship model, with the exception that entities are represented by circles (rather than squares) to be more consistent with traditional mathematical graph theory representation. The graphics are composed of arcs and nodes, with entities types, relationship types, and attribute types all shown as a contiguous graphical representation of the database (Figure 12).

The graphic representation masks the specific values of the data elements to achieve a more general model of the data as does the table format. In addition, it explicitly represents relationships among the data elements, a facility not provided for with the tabular form. The disadvantage to this representation is that the graphical form has little resemblance to an individual data record.

The experiment utilized a single database with enough complex relationships among the elements to make the experiment somewhat challenging to the subjects. The objective was to achieve a balanced degree of difficulty that would yield a significantly broad range of performances to lead to statistical significance of the evaluation. A list of English language queries were developed which required browse-type navigation

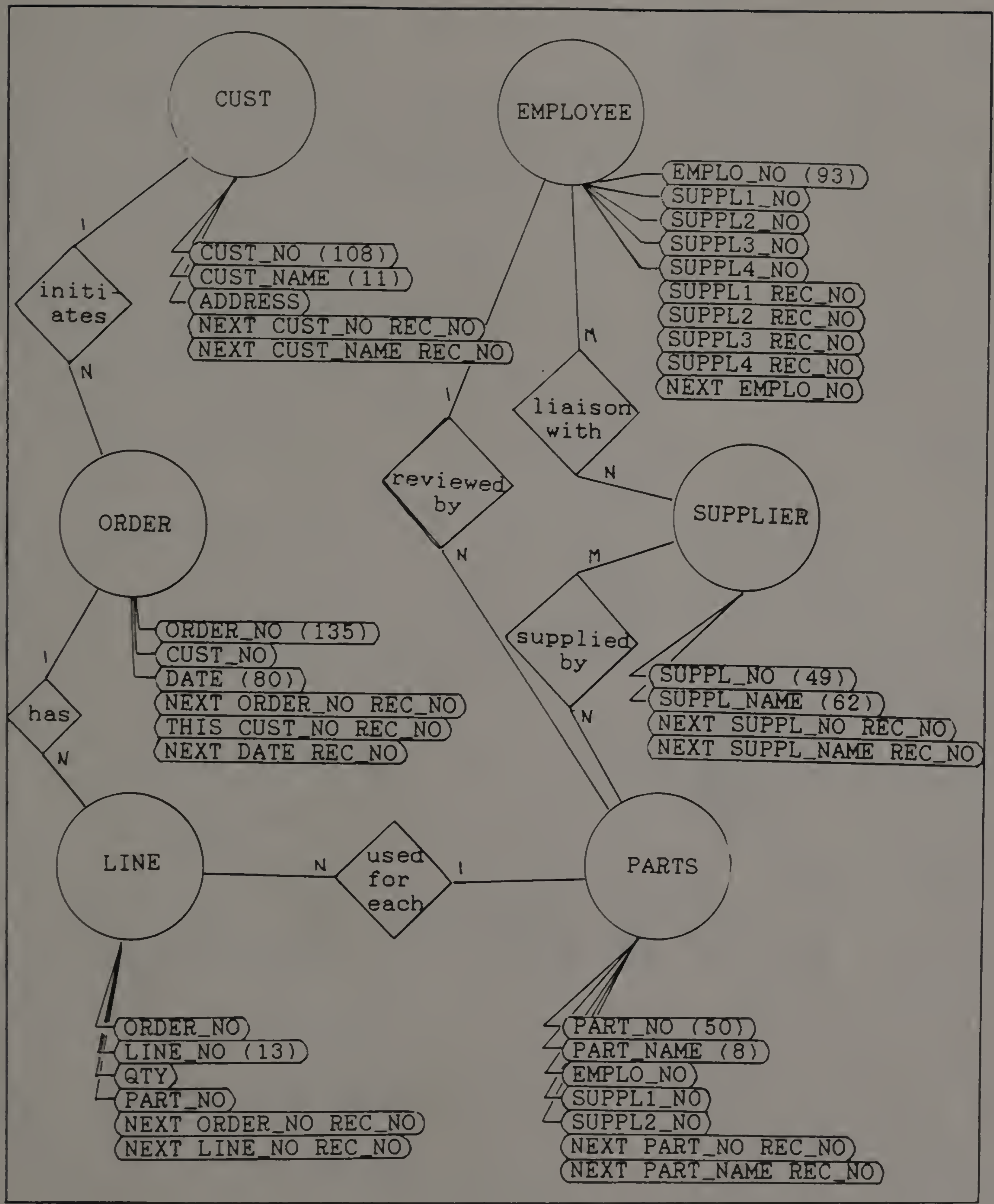


Figure 12 - Sample Graph Documentation

through the database, but a formal query language itself was not introduced. The queries were designed to encourage an iterative search and browse approach, whereby the user viewed individual records of his or her choice, but was prevented from seeing more than one record at a time. The experiment thus required a very crude manual execution of database queries, and roughly simulated the automated query process.

The assumption was that the process can be more efficient and effective when the person performing the queries is able to accurately conceptualize the entire structure of the database. This research suggests that this conceptualization of the database is primarily influenced by the representation of the database and individual differences. The latter factor (individual differences) was diluted through a pre-test grouping process of individuals, plus a second run of the experiment, when users were aided by the opposite form of documentation from their first run.

The experiment drew upon subjects chosen from among reasonably intelligent people without particular sensitivity to their level of computer expertise (e.g. under-graduate students enrolled in an introductory MIS course). This sample group was not considered to be atypical of the larger population of casual users.

The data, or individual records, were viewed on 3x5 cards, arranged by record numbers which were initially assigned randomly, but with all copies of the database identically sorted. Each card displayed a variety of data elements, including further clarification of the data, relationships with other pieces of data, and selected addresses (record numbers, in this case) to facilitate forward navigation through the database.

The quantitative results of the experiment include the time spent by each subject on each query according to the representation type used, as well as the accuracy of the results of the query. The statistical analysis of results regards within-group variations of scores as well as between-group variations, with the perspective of the null hypothesis statement eventually being sustained.

B. The Design of the Experiment

The parameters of the database experiment are herein described as follows. The experiment was conducted as part of an under-graduate MIS class exercise, with the students as captive subjects. The instructor (the researcher) conducted the experiment, after a general discussion about the nature of the exercise. The sample population of thirty students was pre-tested [see also Appendix D], assigned to two homogeneous groups based on the results of the pre-test, and tested again together on two other separate occasions. For data analysis and discussion purposes, they are referred to as the Group A and the Group B. The entire testing process took a total of five hours of contact time in four sessions during a two week period.

There was no Control Group based on the following rationale. The experiment was not of the nature of a treatment versus no treatment, but more the observation of the performance of one type of representation relative to a second type. A control group would suggest that the experiment could be successfully conducted with no data representation at all. In theory, a group provided with no representation, or documentation, would have been feasible, however, their performance would then have to be attributed to factors other than the type of representation and the focus of the study would necessarily be very different than currently envisioned.

A limited amount of demographic data was collected on the subjects themselves. Data was collected on gender, age, database experience, cognitive style, and academic performance [see also Appendix D and E].

The purpose in collecting the data was two-fold. First, it was used to enhance the credibility of the results by showing relative homogeneity of the groups chosen on the basis of a pre-test. This was deemed necessary because of the relatively small size of the sample. The second function was to provide opportunity for further research in the event that correlation between performance and demographic profiles became apparent.

The results of the queries for each subject is the source of the raw data for this research. The number of queries, ten in all, was chosen to provide an exercise to participants that would last about an hour (an estimated attention span prior to mental fatigue). It was assumed (and subsequently established) that ten queries would provide a sufficient range of measurement points to establish the "real" performance of an individual. Lastly, the number of queries chosen allowed different "types" of queries to be tried. This feature presumes that some types of queries may have been more sensitive to representation differences than others. Previous work has suggested that various retrieval techniques exhibit the characteristic of performing better for some types of queries than for others [Belkin and Croft, 1987], [Benbasat, Dexter, and Todd, 1986].

The sample size of fifteen per group was selected in order balance the need for statistical significance with the realities of participant availability. Prior to the actual running of the experiment, there was no available data to suggest how diverse the performance data would be, or in fact if it would be distributed unevenly enough to suggest statistical significance. The ex ante presumption was that if indeed a difference between groups is going to be observable, that it will be revealed with group sizes of fifteen.

A two-by-two-by-two factorial design [Gold, 1984] was used to capture the effects of not only the Graph versus Table phenomenon, but also the effects of doing queries a second time. In the first run of the experiment, Group A used the graphical form of documentation and Group B used the tabular form. With ten queries each, participants were asked to note both the answers to their queries, as well as the time of completion of each query. The time was noted by each subject from a large digital clock accurate to one-tenth of a minute.

Thus, each query has an effectiveness component (based on percentage of right answers) as well as an efficiency component (time elapsed). That is, for the first run, twenty data points (ten effectiveness and ten efficiency measurements) were collected for each of the thirty participants. The experiment was conducted twice, with the groups switching from table to graph forms (and visa versa) to dilute the effects of group difference and to increase the effective sample size. The two runs of the experiment yielded twelve hundred performance data elements which became the basis of the statistical analysis and hypothesis testing.

The rationale in using this particular design was to achieve the primary objective of isolating the effects of users depending on the graph versus the table form of structural representation. In the interest of isolating this factor from the possibility of having differences between the groups, the experiment was run twice, with each group using each form. This "double running" of the experiment had the disadvantage of not only doubling the cost of testing, but also introduced the new factor of whether the test had been taken previously (albeit with a different form of documentation). In spite of these disadvantages, the two-by-two-by-two design (two groups, two representations, two tests) allowed for the most thorough segregation of the effects of the representation factors. Thus, the design provided a classic design and credible experiment with internal validity.

The analysis of the results examined primarily the statistical significance of the differences in the results of the performance of the graph versus table forms. The implications of the results are somewhat tempered by similar analysis of the statistical significance of the results of the group A versus group B results and run one versus run two. Nonetheless, the primary focus remained on the lack of statistical difference between the graph and table performance, and on this basis it was decided that the null hypotheses could not be rejected.

C. The Database

The experimental database contained six major entity types, including Customers, Orders, Lines of the order, Parts, Suppliers, and Employees. In addition, there were six defined links, or relationships, between entity types. The relationship between Customer and Order was called "initiates", as in "the customer initiates the order". Other relationships included "order has lines", "line used for each part", "parts reviewed by employee", "parts supplied by supplier", and "employee liaison with supplier". These relationships were explicit on the Graph, whereas on the Table they were defined by secondary fields on one record type referring to primary fields on a second record type. Each entity had between two and nine purely descriptive attributes. In addition, a few quasi attributes, which were really addresses, by record number, of subsequent records of the same type were listed. These addresses provided the navigational path through the data base.

The complete experimental database is listed in Appendix F as a facsimile of the actual experimental database that was printed on 3 x 5 index cards (one card per record with a total of 135 cards). A simple card holder was supplied to allow easy flipping through the cards. The actual search paths through the database were rather complex, even with only 135 records, as there was no logical order to the cards. This was an intentional characteristic of the database as part of the challenge was to force subjects to think in terms of logical structure and not in terms of the physical arrangement of the records.

The random assortment was attained by first creating a carefully sorted and well organized database in order to establish the queries and the anticipated search paths. Next, the sorted database was repeatedly shuffled (as if it were a deck of playing cards) and record numbers reassigned based on the new random order. The final result was a rather small experimental database which seemed as difficult to navigate through as a very large database. That is, searching for sequential records required using addresses listed on the cards, and could not be guessed or memorized by any of the available subjects.

D. The Queries

The following discussion regards each of the ten test queries. Each query is written as it appears of the test sheet. The Answer and Time blanks for users to fill in are not shown here, but may be seen in Appendix D with a copy of the test instrument. Included also is a Query Analysis, which notes each query's relative degree of difficulty and a description of a search strategy to satisfy it.

The Path Analysis proved a rough measure for ranking the query's difficulty. First the number of different record or entity types (R/E types) that must be browsed is noted. The presumption is that the greater the number of R/E types, the greater the complexity of the problem. In some cases, the number of search paths is noted. Occasionally, a user may be dealing with a single R/E type, but may be required to browse that file on paths indexed on different attributes. This is also presumed to contribute to the complexity of the query. Further, the number of record steps, the total number of cards that must be regarded in order to complete the search, is noted. This number is based on this author's assumptions about what constitutes the most obvious and efficient search paths. In addition, the proposed answer is noted, and lastly, a note about the groups' overall performance on the particular query is given. The performance data is the mean for each group, for each test (designated T1 and T2 in this context), where Effectiveness is "percent correct" and Efficiency is in "net times" to a tenth of a minute (actually inversely related to efficiency).

Query No. 1: Find and list the customer, by name [CUST_NAME], whose name begin with the letter "G".

Query Analysis: This query is intended to be a relatively easy "warm up" query. It uses only a single entity type with a short search path. Users should recognize immediately that this query addresses the record/entity (R/E) type CUST. The first record number, indexed on CUST_NAME is 11. Searches via NEXT CUST_NAME REC_NO yield the following path of record numbers: 11 114 24 31 16 110 48 15.

Path Analysis: (1) R/E type; (8) record steps

Answer: GIBBS

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	1.0	1.0	Effectiveness	.90	.93
Efficiency	14	13	Efficiency	17	14

Query No. 2: Which two SUPPLIERS, by name and number, handle BUTTONS?

Query Analysis: Users begin to practice "backing into" the query with this second query. That is, they must first find the part record identified by PART_NAME BUTTONS, and note the two relevant SUPPL_NOs and corresponding REC_NOs, and then search twice through the supplier list to find their respective names. The search by PART_NAME yields the following path of record numbers: 8 19 133 33 126 83. The final record 83 list the two records of the suppliers of BUTTONS as 62 and 65, which subsequently reveal the suppliers names and numbers.

Path Analysis: (2) R/E types (8) record steps

Answers: #87 FINE, #21 ACE

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.80	.93	Effectiveness	.83	1.0
Efficiency	37	28	Efficiency	44	37

Query No. 3: Which PARTs, by number, were ordered on ORDER_NO 105?

Query Analysis: Progressively more difficult, this query requires users to use a third R/E type not identified in the query (LINE type) to find PART_NOs associated with a particular ORDER_NO. The strategy requires going to the top of the LINE file and moving to each successive ORDER_NO until ORDER_NO 105 is found. This strategy yields the following path: 13 131 86 32. The last record (32) is the first reference to ORDER_NO 105 and reveals the PART_NO associated with the first LINE of the order. Further searches by successive LINE_NOs (of ORDER_NO 105) yield a path as follows: 60 124 68. Each of those records also reveal the PART_NOs associated with its respective LINE.

Path Analysis: (1) R/E type with (2) search paths, (8) record steps.

Answers: PART_NOs 451, 240, 021, 847

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.55	.47	Effectiveness	.53	.68
Efficiency	75	40	Efficiency	74	59

Query No. 4: Employee No. 354 works with which suppliers (by name)?

Query Analysis: This straight-forward query requires a sequential search of the EMPLOYEE file from top of file to the record with EMPLOYEE_NO 354. The path is 93 40 99 25. With this record are the identification of the suppliers and their respective record numbers. The supplier record numbers must then be browsed for their respective supplier names. The supplier search path is 62 57 49.

Path Analysis: (2) R/E types; (7) record steps

Answers: ACE, GOODE, QUALITY

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.93	.96	Effectiveness	.98	.96
Efficiency	37	22	Efficiency	30	31

Query No. 5: What parts (by name) were ordered on ORDER_NO 42?

Query Analysis: This query combines the two path search from a single R/E type as used in Query # 3 with a second R/E type similar to Query #4. Thus the strategy is to start at the top of LINE file and search each NEXT ORDER_NO until the first LINE of ORDER_NO 42 is found. The resultant path is 13 131 86. The last record will show the part number associated with the first line of the order, as well as the record number of that part. Then, NEXT LINE_NO records are successively browsed (records 59 88 14) to determine both the remaining PART_NOs of interest, and also their respective record numbers. These PART record numbers (134 73 89 33) are then directly accessed in turn to determine the PART_NAMEs of the identified parts.

Path Analysis: (2) R/E types with (3) search paths; (10) record steps.

Answers: GASKET, SPACER, NUT, BOLT

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.62	.55	Effectiveness	.47	.82
Efficiency	74	23	Efficiency	61	38

Query No. 6: What is the address of the customer who ordered on the date 07 February?

Query Analysis: While this can be an easy query, it takes a perceptive user to notice information in two separate locations on the physical data record. If the information is overlooked, the search process becomes considerably longer. The first step is to find the order for 07 February via a sequential date search. This path is 80 112 103 102 45. The last record identifies not only the customer by CUST_NO, but at a different location, that customers record number. This record (16) can be accessed directly for the ADDRESS data.

Path Analysis: (2) R/E types; (6) record steps.

Answer: NORTHFIELD

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.97	.87	Effectiveness	.87	1.0
Efficiency	42	22	Efficiency	35	24

Query No. 7: What are the two suppliers (by name) of the part(s) ordered on order no. 112?

Query Analysis: This is a query of few steps, but of greater complexity than previous queries. The strategy is to backward search from ORDER to PART to SUPPLIER. Thus, from the top of the LINE file, a sequential search by ORDER_NO, (13 131 86 32 72) yields the record for the first line of ORDER_NO 112. This record has both the lone PART_NO, and that part's record number (33). The PART record is directly accessed to reveal the suppliers, but by SUPPL_NO only. However, the supplier record numbers are shown (62 57) and these, too, are directly accessed to reveal the SUPPL_NAMEs.

Path Analysis: (3) R/E types; (8) record steps

Answers: ACE, GOODE

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.77	.90	Effectiveness	.97	.83
Efficiency	75	35	Efficiency	64	42

Query No. 8: What part (by name) was ordered in the month of March and who ordered it (by name)?

Query Analysis: This is essentially a double query, as two disparate questions are posed, albeit against the same ORDER. The first search is for the order of the month of MARCH, done by a chronological search (80 112 103 102 45 78 48). This exposes not only the ORDER_NO (647), but also CUST_NO (376) and the customer's record number (48). That record can be directly accessed to yield the CUST_NAME, or half of the double query. Next, the part information is found by searching the LINE records for the previously identified ORDER_NO. The LINE record for ORDER_NO 647 shows the PART_NO (718) and the parts record number (132) which is directly accessed for the PART_NAME.

Path Analysis: (4) R/E types; (29) record steps.

Answers: CAP ordered by DEXTER

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.83	.90	Effectiveness	.73	.83
Efficiency	69	65	Efficiency	103	63

Query No. 9: Is every Supplier assigned to at least one employee?

Query Analysis: Although this is a yes or no question, it is otherwise similar to the other queries in requiring sequential and selective searches of a two files. The strategy for completing the query starts at the top of the SUPPLIER file to identify the complete list of all suppliers by number. The path is 49 62 104 57 65. Next, the Employee file must be searched until each SUPPL_NO is referenced. This is accomplished in the short path (93 40).

Path Analysis: (2) R/E types; (7) record steps.

Answer: YES

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.83	.87	Effectiveness	.87	.73
Efficiency	40	28	Efficiency	42	19

Query No. 10: What parts (by part no.) were ordered in the month of January?

Query Analysis: This is by far the longest query to complete and is moderately complex. "Backing into the search" is required, that is, first identifying the ORDER_NOs for the month of JANUARY using the top of the ORDER file and searching chronologically through JANUARY's ORDERs. The path is (80 112 103 102 45). Next, the individual LINE records must be searched, by ORDER_NO to the first ORDER of JANUARY. The path is (13 131 86). Then, while still working with the LINE file, the search must follow the NEXT LINE_NO path (59 88 14), meanwhile noting the PART_NOs of interest. At the end of LINE file (for that ORDER_NO) lines must continue to be browsed to find the remaining relevant ORDER_NOs. This path reveals two, single LINE ORDERS for which no additional LINE_NO searches need be pursued, and PART_NOs can be noted. The path is (32 72 18 35 38 21 71 6 42 53). The last record (53) is the first LINE of a multiple-line ORDER. Thus, an additional search by LINE_NO is needed to capture the remaining PART_NOs. The path is (100 37 55). This search would have been more efficient with a path through the LINE file indexed by date.

Path Analysis: (2) R/E types; (3) search paths; (24) record steps.

Answers: 625 432 227 53 556 759 546 432 343

Group Performance:

Graph	T1	T2	Table	T1	T2
Effectiveness	.41	.42	Effectiveness	.44	.51
Efficiency	71	83	Efficiency	61	66

CHAPTER 5

DATA ANALYSIS AND CONCLUSIONS

A. Pre-Test

Individual differences in cognitive abilities was assumed to be a significant variable in determining a subject's performance in satisfying the queries of the experiment. However, determining the identity of the specific attributes defining cognitive abilities of an individual was not the objective of the research. Consequently, the following three-question pre-test approach was taken to dilute the effects of those individual differences.

The first ordering of the pre-test results was based on the number of correct responses, with partially correct responses receiving partial ranking points. Scores ranged from all correct to one correct. Within the "number correct" groupings, the "elapsed total time to finish" further discriminated among the students, such that twenty-eight students were able to be ranked on ability. The ranked students were alternately assigned to groups A and B in order to attain parity of ability between the two groups [see also Appendix E]. Two students were absent for the pre-test and were simply assigned to alternate groups.

B. Demographics

The similarity of the two groups can be illustrated by the differences in means of the demographic data of the subjects. The methodology for determining whether there existed statistical significance was a classical hypothesis test for the difference between the means of two populations with independent samples. A two-tailed test, using the t distribution, compared the difference of the means (d) with calculated critical values (d*). The decision rules were as follows.

1) If the absolute value of (d) is less than (d*), then there was no significant difference in the two populations, therefore, the null hypothesis was not rejected.

2) If the absolute value of (d) was greater than (d*), then there was reason to reject the null hypothesis and the alternate hypotheses are then examined.

For initial comparisons a .05 confidence level was used on the theory that differences passing this hurdle could be more stringently analyzed, and differences failing to meet the hurdle could be considered not significant. The author remained sensitive to comparisons that may not have met the hurdle, but may have had a low enough P (probability of being from similar populations) to be worthy of mention.

Relative to gender mix, both groups had identical "male - female" compositions of 11 and 4 respectively. Another significant demographic measurement further indicated relative similarity between groups; that is, the average of the student cumulative grade point average (GPA). On this surrogate measure of intellectual prowess, the means of the two groups varied by only .17 on a 4.00 point scale. Group A's mean GPA was 2.72, and Group B's was 2.55 (comparable to the difference between 82.2 and 80.5 on a 100 point academic scale). Thus it was apparent (and verified statistically) that according to GPA, the two groups were closely comparable.

Two measures of experience were taken for each subject, the first being the age of the individual and the second being a subjective self-assessment as to the person's previous experience working with databases. The age question was posed in such a way as to yield answers with precision only to the whole year, and the difference between the group means was less than one year. Also, subjects were in a range of years from 20 to 24, a rather tight cluster, and the conclusion, statistically confirmed, was that there was no significant difference between the groups on this dimension.

On the second question, self-assessment of previous database experience, one person admitted having some meaningful exposure (this response was valued at 3 points) and the rest admitted having only a limited exposure (value = 2 points), or virtually none (value = 1 point). The groups' means, using the values as previously noted, were 1.73 and 1.8 for Group A and B respectively, and this further supported notions of inter-group homogeneity.

The final demographic measure of interest was based on a construct related to an individuals affinity toward one of two major cognitive styles. One style reflected the logic of the sciences, (the "Math" answer), the other style reflected intuition and the arts (the "Language" answer). Group A responded with an 11 to 4 favoritism of Math over Language, while Group B responded with a 6 to 9 split, with the majority favoring Language. In light of the other demographic measures which indicated group similarity, this "cognitive style affinity" measure appears to identify the greatest difference between the two groups. The significance of this difference was statistically weak ($P < .10$) even with a one-tail test. More importantly, the research significance is difficult to assess because of the softness of the basic construct, and thus may call into question the rationale for including the question at all. The motivation was to include variables that might correlate to performance, and cognitive style, albeit a soft construct, was an intuitively attractive variable to investigate. Its ability to capture a variable with a correlation to performance is shown shortly, but for now, it stands simply as the only notable demographic difference between the two subject groups.

A simple regression matrix was created [see Appendix E] which noted the coefficients of determination, r-squares, for the demographic variables (the independent variables) versus the mean scores of each subject in Test 1 and 2, on performance measures of both efficiency and effectiveness. Specifically, measures of gender, age, database experience, cognitive affinity, and grade point average, were regressed against Test 1 - Efficiency, Test 1 - Effectiveness, Test 2 - Efficiency, and Test 2 - Effectiveness. Within this five by four matrix, only four simple regressions had r-squares over .10, and those were only slightly above. The demographic factors that showed no hint of affecting performance were gender, age, and cognitive affinity.

The factor that began to show some correlation was grade point average, especially in Test 1. As the test of relative strength was having a r-square of greater than .10, the GPA showed strength in both measures of efficiency (r-square = .12) and effectiveness (r-square = .13) in Test 1. It also had strength in Test 2 in efficiency (r-square = .12), but not in effectiveness. The possible explanation for this pattern is that the historically good students had positive attitudes and high motivation on the initial testing, and thus performed better, but that the GPA factor (and the concurrent attitudes and motivations) were less significant in discriminating among students once the entire group became accustom to the querying process.

The other factor worthy of a short mention is that of previous database experience. On Test 1, this factor showed some strength in correlation with effectiveness (r-square = .12). But again, as with GPA the effects on performance in Test 2 were diminished. In general, the conclusion that must ultimately be drawn from the regression matrix is that there was little strong or lasting effect of any of the demographic factors on query performance.

C. Test Results - Efficiency

The two by two by two factorial design allowed comparisons of Group A versus Group B, Test 1 versus Test 2, and most importantly between the Graph documentation and the Table documentation. The measures used for comparison of efficiency were the respective groups' means of query times measured in 1/10th minute time units. The mean for Group A, Test 1, using the Graph, was 53 time units, or 5.3 minutes. For Group B, Test 1, the average time to complete a query was also 5.3 minutes. Given that each group had 15 subjects, and ten queries were completed by most subjects, the number of data points on this measure was almost 150 for each group (actually 147 and 148 for Groups A and B respectively as some students were unable to attempt all ten queries for lack of time). The conclusion is that for Test 1, there was no significant difference between the use of Graphs or Tables.

In Test 2, the groups switched documentation (graph for table), thereby effectively doubling the number of subjects using the graphs and the number using the tables to thirty for each representation format. Besides the advantage of doubling the sample size, the two test approach meant, that by definition, the group that used the tables was made up of exactly the same individuals that used the graph, and thus eliminated the issue of any group / format type interaction.

The disadvantage in having a second test was that it introduced the possibility of interaction between the order of taking the test and the format. That is, it might be argued that using the tables first and the graphs second was materially different than using the graph first and the tables second. A basis for the argument might be that one format is better as initial documentation for the first time users, but for more experienced users (ie. having done the test once before) that the other format was more useful. This question, however, is addressed neither by the hypotheses, by the collected data, or by the experimental design. While the interaction might be considered a threat to the external validity of the experiment, in the final analysis, it was simply presumed not to be a material issue.

More importantly, in Test 2, there began to appear some divergence between results of the graph users and table users. Specifically, the times of the graph users went from the 53 mean time units of Test 1, to 36 mean time units for Test 2, and the table users went from 53 to 39 mean time units. For both groups, the mean time units to complete a query were reduced significantly, but the critical data lies in the final difference of 36 and 39. By definition, the relationship between time to process a query, and the construct of performance efficiency is inverse. Thus, as the group using graphs in Test 2 had lower mean times, their performance was more efficient. The significance of this difference in mean scores was tested in light of the formal hypotheses, where the null hypothesis was that there was no statistically significant difference.

Yet the differences were found to be not statistically significant. Therefore the null hypothesis could not be rejected for any overall difference in the use of graphs versus tables in terms of efficiency of performance.

D. Test Results - Effectiveness

The construct of effectiveness was measured with scores of percent correct answers retrieved for each query. No penalty was assessed for wrong answers; they were simply ignored. Ten graded queries for each of the thirty subjects yielded 300 data points for each test. The summary results were in terms of independent mean scores for each subject in each group, and tests of significance were applied against these subjects' mean scores as in the case of efficiency measures.

The pattern of results showed means slightly higher for the Graph users in Test 1 (Graph = 77.1, Table = 75.8) and slightly higher for Table users in Test 2 (Graph = 78.6, Table = 83.0). These results would have been challenging to explain had there been statistical difference, but as the null hypothesis was not rejected for either test (no difference between formats) the reversal of dominance needs no explanation, as one representation form was as likely as the other to prevail.

It is worthy of explicitly stating that the combination of the two tests bring the results even closer together. The mean scores for the two tests for effectiveness for the Graph users was 77.9 and for the Table users 79.4. These aggregate means were also tested and found not to be statistically different.

A comment about the possibility of the Graph / Table factor having a different effect on different query types is in order. Previously, in the query analyses, the relative difficulty of the various queries was noted. Investigation of the data, however, showed that the relative dispersion of scores between user groups was modest, and on the more difficult queries, there existed a reversal in performance dominance between the Graph and Table users between tests. The conclusion was that the Graph / Table factors influence on performance was not affected by query difficulty in this experiment. The overall results can be summarized in the following statement: Regarding measures of performance for both efficiency and effectiveness, no statistically significant difference was found between users of the Graph form of representation and the Table form.

E. Threats to Internal Validity

Beyond the actual numbers gleaned from the experiments, it is important to acknowledge the threats to both the internal validity of the experiment as well as the external validity, or ability to draw conclusions from the results.

In regards to internal validity, similarities in the overall experimental design to previous tests that have been considered rigorous [Greenblatt and Waxman, 1978], [Benbasat and Dexter, 1985] give credibility to the approach of this research. The preliminary orientation lecture to the students was done in a batch mode, such that all subjects received similar pre-test treatment. The pre-test evaluations and subsequent two phases of the experiment were also completed in batches. For two cases (out of thirty) for the pre-test, and for four cases (also out of thirty) for Test 2, subjects who missed the scheduled test were re-tested during a make-up period. These make-up tests were considered materially similar to the original tests.

The classroom environment of the test provided a purity to the testing process and to the test grading. Further, the query timing process was considered appropriately consistent for all subjects. There did exist the possibility of subject cheating, that is, copying query answers from a neighbor's answer sheet, however, none was observed. A subjective assessment of the environment was that the subjects conducted themselves honorably.

The last notable internal threat involved subject failure to record query completion time in a few instances. In these cases, interpolation of intermediate times were made, with the conclusion that, while "tainted", the data still reflected a reasonable measure of time allocation.

F. Threats to External Validity

In regards to threats to external validity, the assessment is necessarily more subjective. An ex ante concern about the formats of the graph and the tables lies in their ultimate similarity of informational content. In the design of the experiment, there was a fundamental conflict, not unlike with Lochovsky's testing of data models and data languages [Lochovsky, 1978]. That is, in order to test one model versus a second model, it was necessary to simultaneously test peripheral attributes inexorably linked to the their respective models. In Lochovsky's test, the peripherals were in the data language; in this research they were the model primitives, or structural elements.

The resolution of that conflict in this research entailed keeping the various primitives to a minimum in order to focus on the basic representational forms. In the end, the two representations were notably similar. Had the Lochovsky approach been applied to this research, so many more structural elements could have been added to both representations that the essence of testing the difference in the fundamental representation would have been effectively obscured.

An ex post concern deals with the reliance of subjects upon the documentation to formulate their individual query strategies. Informal comments by subjects after the tests were initially discounted due to the inability to analyze their content with scientific rigor. Nonetheless, one comment suggested that for the easy queries, the notion of a search strategy was discarded and the subjects basically bludgeoned their way through the database to find answers. They admitted this technique was futile with more difficult queries.

Also, for these more difficult queries, the author questioned to what extent subjects were relying upon the representational documentation to formulate their search strategies, versus relying upon previous knowledge of the database structure (gathered from previous searches) to navigate through the records. There were no provisions made to measure this phenomenon, nor was it possible to assess the extent of its significance. Hence, it remains simply a noted threat to external validity.

The aforementioned threats to validity are offered in the spirit of candor, but are not intended to undermine the overall significance of the experiment. Authorship bias and noted threats notwithstanding, the results are still considered to have been gleaned from an experiment with sufficiently stringent rigor to yield conclusions, that are, in good faith, fully defensible.

G. Conclusions

The result of this research is that no statistically significant difference existed between the performance of two groups querying a database with different conceptual design representations. The groups were arguably representative of reasonably intelligent, albeit relatively unsophisticated database users. The database proved to be sufficiently complex to be challenging to navigate within, while being sufficiently finite to allow some limited, but measurable success. The classical two-by-two-by-two experimental design and the purity of the test environment both lent integrity to the empirical data.

In its embryonic stages, this work sought to challenge the existing tabular standard used in database query languages. Others have noted that in order to become a significant challenge to the Relational model, the E-R, or graphical approach, would have to take full advantage of the rich data structures expressed by the E-R model and its extensions [Bodart, 1987]. While this experimental database, and the accompanying queries, were not designed to fully express this rich data structure, the experiment clearly compared the essential visual format differences of the two representations.

As the research continued however, it became apparent that there existed very little theoretical evidence to suggest that one form would lead to better performance than the other. Indeed, personal preference and individual differences appear to be more significant factors in selecting a format than any performance considerations.

The significance of this observation is that, by virtue of establishing that neither type of representation of structure leads to more efficient and effective searches of the database by novice users, the following inferences might be made:

1) Database designers should consider several forms of representation of data structures to appeal to a broad user population with varied tastes. This might increase the effective and efficient use of the database as more users could find a representation form that appeals to them personally.

2) Users, when defining their information needs, might also consider more than one representation of their particular data structures as this might increase their ability to express their perceived needs.

A multiple perspective approach to data representation could still offer a common set of representations of data structure to users, analysts, designers, and programmers.

The idea of providing multiple representations for a database has implications for future research. For example, if the representations are available as an on-line help screen, then provisions for tracking frequency of the calls to the different forms could be included. Systems analysts could determine which forms were most popular, more performance data could be collected, and other formats (or refinements) could be offered and tested.

As the more general objective of this work is to contribute to the knowledge of what constitutes quality in database design, it is appropriate to express that beyond the classical measures of effectiveness and efficiency, that one of the significant attributes contributing to the quality of a database, is user satisfaction [Nolan, 1974]. As user satisfaction is likely to be a partial function of the intuitive appeal of the representation of the database then this multiple user approach could lead to greater user satisfaction.

In conclusion, the experiment conducted here suggests that the conceptual design representations currently used to describe real world databases are selected, not by real measures of performance, but by other factors. Those factors undoubtedly include personal taste, and probably include others such as institutional requirements, or system constraints. The results of this research refutes empty rhetoric which says that one form is better than another in terms of performance, and refocuses the issue on multiple conceptual design representations for multiple users.

APPENDIX A

Database Development Process

The continuum of the database design process in which data models are used runs from the real world phenomenon on one end to machine code at the other. This process of building an abstract representation of the enterprise includes the following four phases described below [Tsichritzis and Klug, 1978].

In Phase 1, people most closely associated with the rules of the enterprise define and explain the objects (entities) and their relationships to an analyst creating the enterprise model.

In Phase 2, these verbalized entities are formalized, given structure, embellished with attributes, and given interconnections (relationships) with other entities of the enterprise. The product of this formalization process is the conceptual data model of the enterprise. It presumably defines all the entities of interest and provides a foundation for the actual definition of the data for the database management system (DBMS).

Creating the database via a DBMS is Phase 3. During this phase, the data is further defined, and the enterprise rules are translated, or mapped to a DBMS.

In most cases the DBMS is a commercial package and is one of the three classical types, Hierarchical [IBM, 1975], Network [CODASYL, 1971], or Relational [Codd, 1970].

Through a data definition language (DDL), the entities and relationships from Phase 2, are coded into the system, thus completing Phase 3 and simultaneously activating Phase 4, the actual writing of the data to a physical medium.

It is through these four phases of the design process that the enterprise is modeled. Although the entire process or phases of the process may be iterative in nature, and overlapping in time, there is a discernible difference between the phases. That is, the results of one phase does not absolutely determine the following phases.

The relative independence of the phases suggests, for example, that the same enterprise may be modeled differently, and that it is possible that the same conceptual model may be mapped to any of the three DBMS types (although not always). In addition, the same conceptual model may be represented differently without affecting the model, and lastly, the same DBMS may run with different physical configurations.

APPENDIX B

Model Attributes

The preceding appendix defined conceptual data models in terms of their relative position in the design process. Equally important for understanding the nature of conceptual data models are the qualitative attributes of the models themselves. They may be compared on their expressiveness, naturalness, implementability, and simplicity [Lum, 1978].

Expressiveness: In this context, expressiveness, or power, is the ability of the model to communicate the often subtle, and complex, entities and relationships of the real world. Conversely, overly crude, or constrained models are those which lack the necessary features to fully express essential elements of a real world enterprise.

Naturalness: The term naturalness denotes the quality of expressing the real world elements in a manner that has intuitive appeal. Creators of several data models seem especially pleased with the naturalness of their particular model [Shipman, 1981]. They can claim this with relative impunity as their own models are presumably a natural extension of their own view of reality, and as naturalness is so difficult to otherwise measure.

Implementability: This attribute defines how effectively the conceptual model can be mapped to the later phases. While it was previously noted that implementability was an issue for first generation models, there is a distinction according to the context in which the term is used. That is, a conceptual data model should be able to be mapped to a DBMS with most features intact, while a DBMS must be able to be implemented on a working physical system, and its performance must be competitive.

Simplicity: The last term, simplicity, is the quality of performing all of the above without being encumbered with an unwieldy set of working elements. These are, indeed, the "elegant" models [Chen 1980].

Schemas: The results of conceptual data models that have the above qualities are the specific conceptual designs, or schemas, which will similarly exhibit the following attributes [Lum, 1978]:

1) Minimality: allows for the extensive expression of meaning with only a few symbols,
2) Completeness of thought: acknowledges that closure can be achieved when all elements of a thought are expressed,

3) Evolvability: the quality of the model's ability to change with the enterprise,

4) Correctness: an expression of accuracy,

5) Modularity: allows a conceptual model to be connected with (or disconnected from) other modular designs,

6) Implementability: again, is the ability to be mapped to a DBMS.

APPENDIX C

Selected Models

The following discussion of particular data models proposed over the last twenty years also includes the first generation models (hierarchical, network, and relational) as a reference foundation. These were the precursors of the second generation models, and thus are not considered purely conceptual data models. Their orientation is more toward the implementation phase of database development and this emphasis is often at the expense of their being able to capture complex relationships of the real world.

The Hierarchical Model: One of the oldest data models is IBM's Information Management System (IMS) [Tsichritzis and Lochovsky, 1982]. The earlier discussion of data models in this paper noted that the implementation of this structure actually preceded the structure being defined as a "data model".

The structural representation of a hierarchical database is a graph in the form of an ordered tree (Figure 13). The general model of an ordered tree is one in which the relative position (both vertical and horizontal) of nodes (or leaves) carries semantic meaning. Further, the arcs have direction, pointing toward the nodes and may be referred to as hierarchical definition trees [Tsichritzis and Lochovsky, 1982].

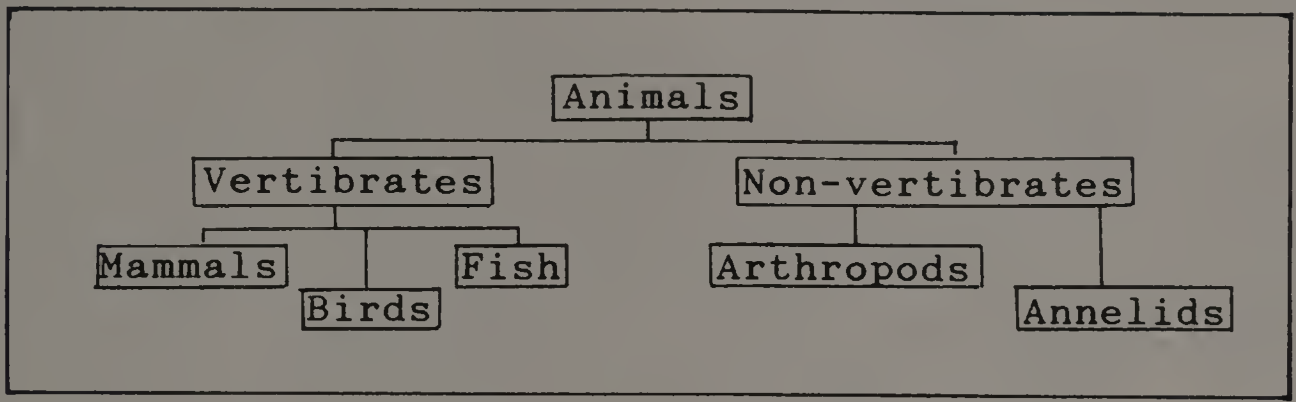


Figure 13 - Hierarchical Structure

The arcs are functional links called "parent-child relationships", and there exists the general restriction that each entity have one and only one parent. "Functional", in this context, means that a record can have at most one parent of any record type.

"Ancestor" and "descendent" entities may be identified by following arc paths up or down the hierarchy. By definition, the "root" entity has no parent, and contains a single record type. [Tsichritzis and Lochovsky, 1982].

The Network Model: Network databases, like hierarchical databases, had instances of fully operational applications prior to being formally defined as data models. This definition came in the form of a report by the Data Base Task Group (DBTG) of the Conference on Data Systems Language (CODASYL) [CODASYL, 1971].

As a structure for a data model, the Network configuration is defined in terms of tables and graphs, where the nodes on Bachman diagrams represent the entity types, which are also represented by tables. The arcs represent the relationships between the entities as with the hierarchical model, but the constraints related to the arcs are significantly different (Figure 14).

In the simple network data model the functional nature of the arcs prevents a member from having more than one owner, thus excluding the facility to define many-to-many relationships [McFadden and Hoffer, 1985]. Real world situations which would otherwise suggest a many-to-many type of relationship may be expressed with the addition of intermediate nodes. The simple network model is thus able to achieve a necessary level of expressiveness, while maintaining a rigorous set of operating constraints.

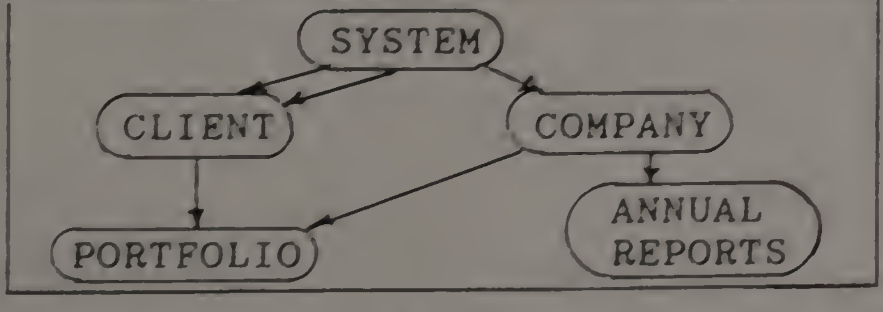


Figure 14 - Network Structure

The Relational Model: Codd proposed the Relational data model based on mathematical set theory in 1970 [Codd, 1970]. In the model, mathematical sets (or relations) were represented by tables (Figure 15), and the data contained on those tables were subject to the logic of relational mathematics. The model has obvious advantages on some dimensions, specifically the ability to maintain data integrity, due in part to the application of Normal Form Theory. In other areas, such as performance and expressiveness, the agreement on its advantages is less decisive.

CUSTOMER

NO	NAME	ADDRESS
036	Chen	Norfolk
376	Dexter	York
709	Kent	London

ORDER

NO	CUST_NO	DATE
001	404	29 Mar
002	036	14 Apr
003	709	13 Jun

Figure 15 - Relational Structure

For example, data models based on the original Relational model do not provide the semantic richness found in other models. This is due, in part, to the obfuscation of the associations between entities that occurs in the normalization process. Further, it is generally necessary to impose a single structural organization on the data, thus defeating the potential for supporting relativism. These issues were subsequently addressed in Codd's Extended Relational

Model RM/T [Codd, 1979], which is discussed later in this paper.

The structural elements of the original Relational model, such as relations (tables), tuples (corresponding to records or rows on a table), and keys(attributes or fields) are perhaps less critical to elaborate upon, as the original model is not widely accepted as one of particularly impressive semantic expressiveness.

The Binary Data Model: A Binary Data Model is a graph data model in which nodes represent simple single attributes, and the arcs represent binary relationship type between the two attributes [Tsichritzis and Lochovsky, 1982]. One of the first proposed binary data models, and one of the earliest second generation models was the Semantic Binary Data Model [Abrial, 1974].

Abrial tests the power of his model by using it to describe itself, with the implication that if a data model can be used as a metamodel to describe itself, it can also be used to describe other data models. Using a precise metamodel to describe different data models would provide a common framework for understanding and comparing data models [Abrial, 1974].

The Basic Semantic Model: Another of the early second generation models is the Basic Semantic Model [Schmid and Swenson, 1975]. This model takes the perspective that the real world consists of objects (entities) and associations (relationships). Their fundamental data construct is the n-ary relation, or relationship among two or more entities.

In this model, objects can be existence dependent on other objects, or can be independent, and further, objects can have descriptive characteristics, or properties [Date, 1985]. Relations are classified according to the type of information they represent. This classification scheme was seen as a basis for relativism, or the facility of providing multiple views of a database at multiple levels of abstraction [Date, 1985].

The Entity-Relationship Model: Of the various conceptual models that have been proposed, the one that appears to have attracted the most attention is Chen's Entity-Relationship Model (E-R model, or ERM) [Chen 1976]. According to the objectives of the model, an enterprise schema created via the E-R model will be both independent of, and stable relative to, changes to the host DBMS.

Thus, the schema represents the conceptual structure and nature of the data of the enterprise without preoccupation with the actual implementation of the model. Chen's original model provides both a foundation of simplicity and a point of departure for models of greater complexity.

The primary elements of the model are the entities (the objects, the things) and relationships, which describe the generic structure among the entities. The entities may be "weak", that is, existent dependent on other entities, or "regular" (not existence dependent).

Both types of dependencies may exist in the same model, and they are clearly expressed by one of several different graphical conventions (for example see Figure 10, p. 29). With existence dependencies, Chen assumed that the deletion of an entity would "cascade", that is, would cause the subsequent deletion of dependent relationships and entities.

Entity-Relationship diagrams clearly provide for the representation of one-to-many (1:N) and many-to-many (M:N) relationships between entities, or cardinality, as shown below (Figure 16).

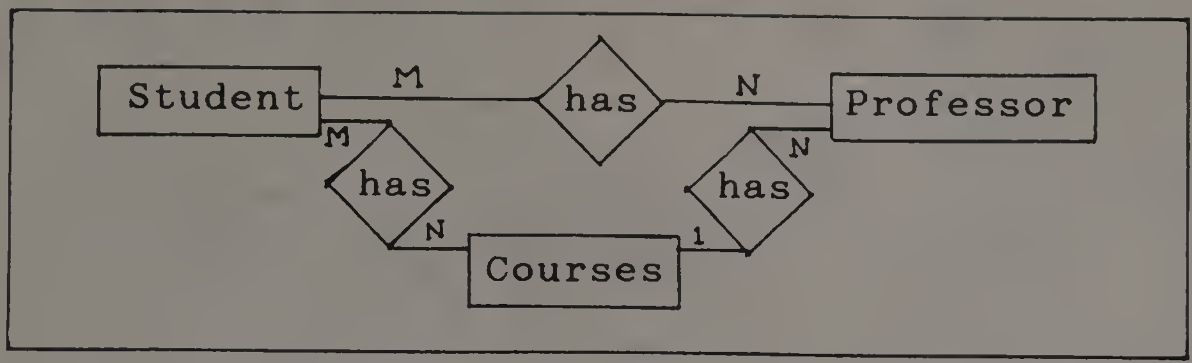


Figure 16 - Cardinality in the E-R model

In most variations of the E-R model, the concept of "entity" remains unaltered as the atomic expression of a thing. The differences among the variations in the E-R models lie in the interpretation and constraints on the concepts of "relationships" or "attributes" [Chen, 1981]. Based on these different treatments of relationships and attributes, Chen proposed the following classification scheme to fit various extensions of the E-R model.

In terms of relationships, models are differentiated by whether they allow n-ary relationships (more than two entities connected to a relationship) or allow only binary relationships.

A binary relation is intended to represent a single type of "atomic fact", or a relationship between a single entity type and a single corresponding property type [Date, 1985]. N-ary relationships, on the other hand allow for multiple relationships, either one-to-many or many-to-many, thus providing considerably more expressiveness at the cost of greater complexity. Dividing the models on this basis yields Generalized Entity Relationship Models (GERMs) and Binary Entity Relationship Models (BERMs).

In terms of attributes, the models are differentiated according to three conditions. They are, first, whether attributes are allowed for both entities and relationships, second, whether attributes are allowed for entities only, or lastly, whether attributes are not allowed at all. In cases where attributes are not allowed, the attribute function is provided by additional entities and relationships.

Somewhat of an orphan to this framework, but related historically, is the TERM, a time oriented E-R model [Klopprogge, 1981]. TERM allows for the definition of application specific patterns of temporal change, including concepts for the handling of incomplete or imprecise data, correcting erroneous data, and maintaining continuous histories. More simply, it adds the attribute of time to entities, relationships, and attributes, thus providing for greater expressiveness.

Other references to the time dimension include a paper by Schueler advocating the outlawing of deleting and destructive overwriting [Schueler, 1977], an implementation of Schueler's concept [Jones and Mason, 1980], and other time oriented proposals [Bubenko, 1977], [Robinson, 1979].

The Semantic Database Models: The term "Semantic Data Model" refers to both a specific model [Hammer and McLeod, 1981] as well as a set of data models [Date, 1985]. Within the set of Semantic Data Models is not only the Semantic Data Model (SDM) of Hammer and McLeod, but also an earlier Basic SDM [Schmid and Swenson, 1975], DAPLEX [Shipman, 1981], and a hierarchical type model [Smith and Smith, 1977]. These models are discussed separately.

The Semantic Data Model: Hammer and McLeod designed their SDM with the intention of capturing "more of the real world environment" than they thought was previously possible with existing data models [Hammer and McLeod, 1981]. Their approach was to improve upon the existing hierarchical, network, and relational models, by adding facilities that were otherwise lacking.

In the design of SDM, the designers sought a balance between an "absolute minimal number of mutually orthogonal constructs", and a profusion of special case facilities. They noted the significant trade-off between the complexity of a modeling facility and its power, naturalness, and precision.

SDM is based on several general principles of data organization as noted here.

1) The model is a collection of entities,

2) Entities are organized into classes,

3) Classes are logically related by virtue of interclass connections,

4) Entities and classes have attributes, which may be derived from other values. These attributes describe characteristics and define relationships,

5) Several primitive ways of defining interclass connections and derived attributes exist. The facilities integrate multiple ways of viewing the same basic data, and provide building blocks for describing complex attributes and interclass relationships.

DAPLEX: While actually a data definition language (DDL), DAPLEX [Shipman, 1981] also fulfills the role of a conceptual data model. It incorporates the concept of entities, a functional representation for both virtual and actual data relationships, a collection of language constructs, and a notation of subtype and supertype relationships.

DAPLEX is a functional model, also described as a "directed binary relational model", in which entities are unary relations and functions are represented by directed, many-to-one binary relations. It is the "syntactic embodiment" of the Functional Data Model [Date, 1985] and incorporates notions of semantic data nets [Quillian, 1966], derived data [Hammer and McLeod, 1981], and aggregation and generalization [Smith and Smith, 1977]. The construct of polymorphism is also described whereby a single entity may be defined as being of more than one type.

Data Abstraction Model: The abstraction process allows masking irrelevant details to enhance the upper level structure of the model. A major contribution to the accepted approaches to abstraction are developed in the Data Abstraction Model [Smith and Smith, 1977].

They propose the constructs of aggregation and generalization as enhancements to the conceptual model building process. These facilities have become incorporated in many subsequent data models and thus they have achieved permanence as model primitives.

Type hierarchies, IS_A hierarchies, and PART_OF hierarchies can all be traced to their model.

The Role Model: The Role Model [Bachman and Daya, 1977] was motivated by the ANSI/X3/SPARC specifications and is designed to allow single entities to play different roles in the same conceptual schema. The concept of the record to represent the existence of an entity is retained, where the concept of role segment to represent the existence of one of the entity's roles is introduced.

Role is a defined behavior pattern which may be assumed by entities of different kinds. Further, a given entity may simultaneously play one or more roles. The term role is also used with the Relational model [Codd, 1970] to describe the way in which an attribute relates to a domain. In the E-R model, it describes the way that an entity relates to a relationship [Chen, 1976], and in the Object-Role Model [Falkenberg, 1976] it is used similarly to the original Bachman Role Model.

The Extended Relational Model RM/T: Codd's response to criticism of the original Relational model as lacking in semantic expressiveness is found in his Extended Relational Model RM/T [Codd, 1979]. The model is based on entities, properties of the entities, associations between entities, subtypes and supertypes, and surrogates.

Entities are classified as either characteristic (a one-to-one representation of another entity), as associative (many-to-many relationships), or as kernels (neither characteristic nor associative) [Date, 1985]. RM/T has suggestions for addressing the time dimension in databases, as well as relativism (or the support of alternate views of data). Aggregation is recognized as either Cartesian in nature (an entity as the sum of its parts), or as cover aggregation (entity as a set of heterogeneous members) [Date, 1985].

Current Developments: In recent years, there appears to be more concentration on further refinements of existing models than on radically new data models.

An example is the data model of ECRINS/86 [Junet, 1986] which extends Chen's original E-R model to include not only the abstractions of Smith and Smith, but also the facility to define relationships between relationships. Further, the concept of a role in a relationship is extended to a multi-valued role, which enables the association of a set of tuples performing the same function in a relationship.

Object-orientation: In addition, object-oriented models are receiving a good deal of current attention. Data models with object-orientation are based on the notion that an entity, whatever its complexity and structure, may be represented by exactly one object (or entity) [Dittrich, 1986]. Artificial decomposition into simpler concepts, as in normalization, is not necessary.

Objects may have operations associated with them, such that the reference to a particular object would also be, by definition, a reference to the inherent operations of the entity.

APPENDIX D

Test Instrument

The following pages are copies of the actual written material provided to the subjects of the database experiment. The material include:

Demographic questions 109

Rules for executing queries, and

Sample documentation 110

A verbal description of the database 111

Table documentation 112

Graphic documentation 113

The pre-test 114

The queries 115

The Test Instructions:

You are about to participate in a research experiment being conducted as part of a PhD dissertation to be submitted to the School of Management, University of Massachusetts at Amherst. It is important to the researcher that each participant attempts to respond to the questions in good faith and with their best effort. Results of the experiment will be made available to participants upon request.

Some demographic data on you, the participants, may add insight into the significance of the findings. This information will be kept confidential as required by law. To ensure this confidentiality, this information will in no way be connected to individuals by name.

1. You are... [] male
[] female

2. Your age is.. [] 18 [] 22
[] 19 [] 23
[] 20 [] 24
[] 21 [] over 24

3. You might classify your total experience working with databases as...
[] virtually none
[] a limitted exposure
[] some meaningful exposure
[] extensive experience

4. Two groups of individuals are described with the following terms. With which group would you feel it most appropriate to include yourself?

[] group 1	[] group 2
Mathematics	Languages
Logic	Intuition
Science	Art
Objects	People

5. What is your current overall grade point average (GPA)? (Use a best guess if not precisely known)

GPA = __________

Thank you for your participation.

Edward Harding
PhD Candidate
School of Mgt., UMass

Test Instructions: page 2

Here are some specific instructions for working through this database query exercise:

1. Work on each query in order. Make a best effort to finish each query before moving on to the next. It is important NOT to skip around trying different queries. In general the queries get progressively more and more difficult.

2. Mark the time that you finish working on each query in the space provided. If you are unable to formulate a query search, or you eventually give up trying to complete the query, note the time that you aborted your attempt, then go to the next query.

3. A few queries have more answer blanks than true answers. Use what you need and ignore the rest.

4. If and when you finish working on all 10 queries, you may stop the test.

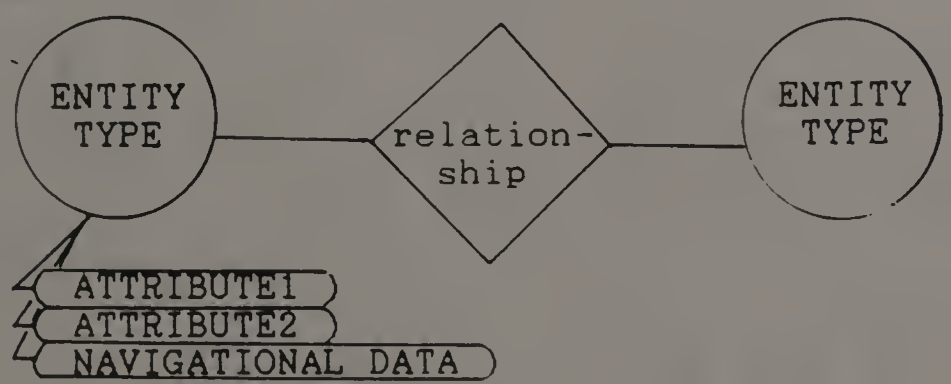
5. The format for both the TABLE representation and the GRAPH representation are shown below, with a description of the type of data found in each location.

6. Note that TOF represents the record number (REC_NO) of the first record of a particular record type, indexed on the field or attribute in which it appears. Also note that navigational data are included as fields or attributes, although they are only descriptive of the parent record or entity in a very loose sense.

Table format:

```
RECORD TYPE
PRIMARY FIELD (TOF) ¦ FIELD1 ¦ FIELD2 ¦
====================¦--------¦--------¦
NAVIGATIONAL DATA ¦
- - - - - - - - - - ¦
```

Graph format:



A Verbal Description of the Database:

This database reflects the activities of a simplified enterprise. It comprises six related files (or record types, or entity types), one for customers, orders, lines on the orders, employees, suppliers, and parts.

The relationship between these different types is described as follows: Customers initiate orders in the process of making a purchase. A single customer may order more than once, but each order may have only one customer. Each order has one or more lines, so that a separate line may be used for each different part. Parts are supplied to the enterprise by outside suppliers. Each part has a primary and secondary supplier, and each supplier may provide several different parts. Each employee has a responsibility for reviewing the stocking levels of certain parts. When the stock level get low on a particular part, the employee contacts one of the two suppliers of the part for replenishment. Thus, each employee has a liaison with several different suppliers.

Customers (CUST) are each assigned customer numbers (CUST_NO). Each customer also has a name (CUST_NAME) and address (ADDRESS). The customer file may be searched by CUST_NO or CUST_NAME.

Orders (ORDER) have order numbers (ORDER_NO), customer numbers (CUST_NO) and order dates (DATE). Orders may be searched by ORDER_NO or by DATE.

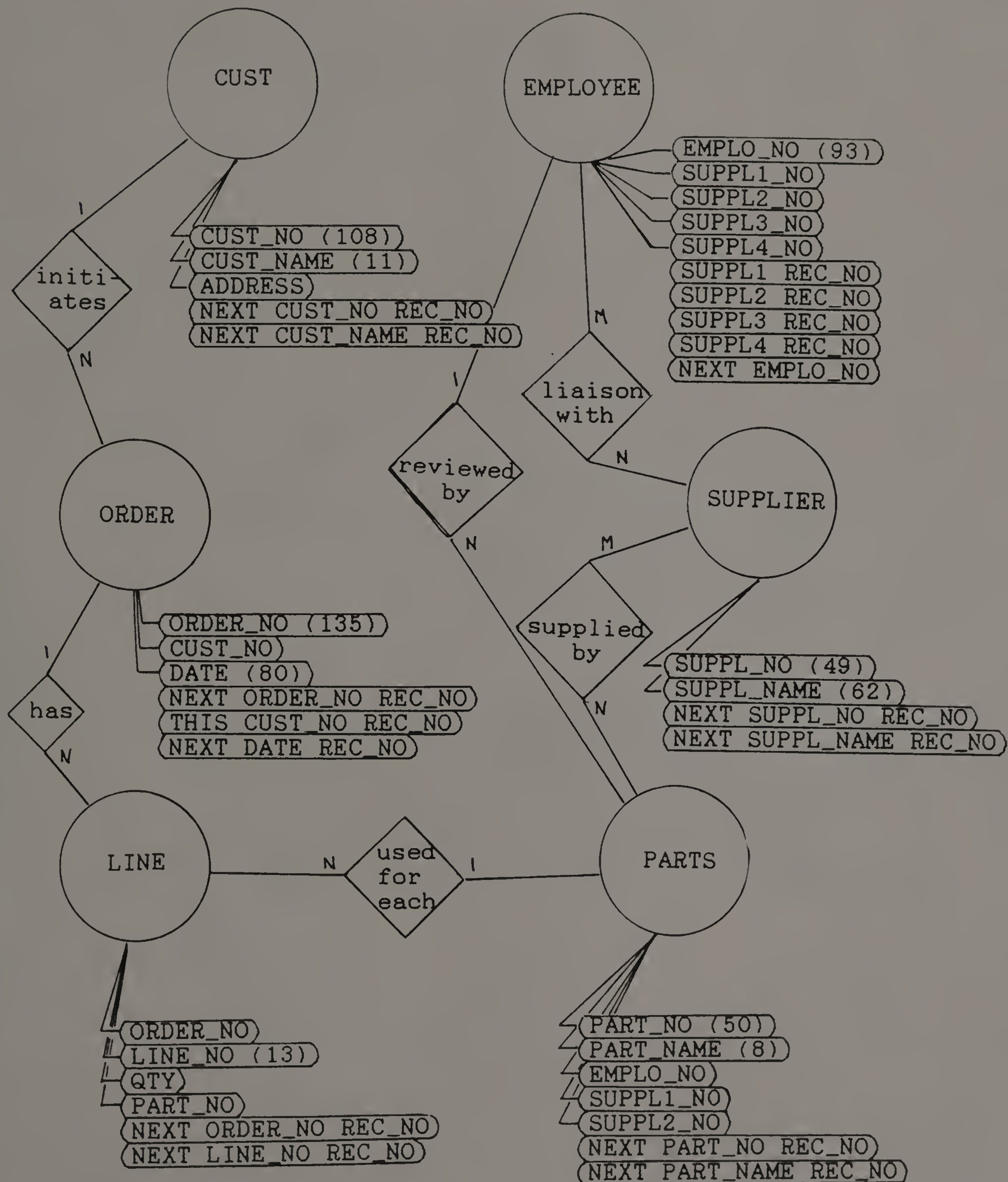
Lines of each order need both the ORDER_NO and line number (LINE_NO) to fully identify it. Lines also show quantity of parts ordered (QTY) and the related part number (PART_NO). The line file may be searched by regarding the first line of consecutive orders or by regarding consecutive lines of the same order.

Parts have names (PART_NAME) and part numbers (PART_NO) and may be search on either characteristic. Also noted is the employee number (EMPLO_NO) of the person responsible for sufficient stocking levels, and the supplier numbers of the primary and secondary suppliers (SUPPL1_NO) and (SUPPL2_NO).

Suppliers are identified by number (SUPPL_NO) as noted above, as well as by name (SUPPL_NAME). The suppliers may be searched on either characteristic.

Employees are identified by number (EMPLO_NO), but not by name. Also noted are the suppliers with whom the employee has liaisons (SUPPL1_NO), (SUPPL2_NO), SUPPL3_NO), and (SUPPL4_NO).

```
CUST
CUST_NO (108) ¦ CUST_NAME (11) ¦ ADDRESS ¦
==============¦----------------¦---------¦
NEXT CUST_NO REC_NO ¦ NEXT CUST_NAME REC_NO ¦
- - - - - - - - - - ¦- - - - - - - - - - - -¦


EMPLOYEE
EMPLO_NO (93) ¦ SUPPL1_NO ¦ SUPPL2_NO ¦ SUPPL3_NO ¦ SUPPL4_NO ¦
==============¦-----------¦-----------¦-----------¦-----------¦
SUPPL1 REC_NO ¦ SUPPL2 REC_NO ¦ SUPPL3 REC_NO ¦ SUPPL4 REC_NO ¦
- - - - - - - ¦ - - - - - - - ¦ - - - - - - - ¦ - - - - - - - ¦
NEXT EMPLO_NO ¦
- - - - - - - ¦


ORDER
ORDER_NO (135) ¦ CUST_NO ¦ DATE (80) ¦
===============¦---------¦-----------¦
NEXT ORDER_NO REC_NO ¦ THIS CUST_NO REC_NO ¦ NEXT DATE REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - ¦ - - - - - - - - -¦


LINE
ORDER_NO ¦ LINE_NO (13)  ¦ QTY ¦ PART_NO ¦
=========¦===============¦-----¦---------¦
NEXT ORDER_NO REC_NO ¦ NEXT LINE_NO REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - ¦


SUPPLIER
SUPPL_NO (49) ¦ SUPPL_NAME (62) ¦
==============¦-----------------¦
NEXT SUPPL_NO REC_NO ¦ NEXT SUPPL_NAME REC_NO ¦
- - - - - - - - - - -¦ - - - - - - - - - - - -¦


PARTS
PART_NO (50) ¦ PART_NAME (8) ¦ EMPLO_NO ¦ SUPPL1_NO ¦ SUPPL2_NO ¦
=============¦---------------¦----------¦-----------¦-----------¦
NEXT PART_NO REC_NO ¦ NEXT PART_NAME REC_NO ¦
- - - - - - - - - - ¦ - - - - - - - - - - - ¦
```



CUST
EMPLOYEE
EMPLO_NO (93)
SUPPL1_NO
SUPPL2_NO
SUPPL3_NO
SUPPL4_NO
SUPPL1 REC_NO
SUPPL2 REC_NO
SUPPL3 REC_NO
SUPPL4 REC_NO
NEXT EMPLO_NO
CUST_NO (108)
CUST_NAME (11)
ADDRESS
NEXT CUST_NO REC_NO
NEXT CUST_NAME REC_NO
initi-
ates
1
N
M
liaison
with
N
1
reviewed
by
N
ORDER
SUPPLIER
M
supplied
by
N
ORDER_NO (135)
CUST_NO
DATE (80)
NEXT ORDER_NO REC_NO
THIS CUST_NO REC_NO
NEXT DATE REC_NO
SUPPL_NO (49)
SUPPL_NAME (62)
NEXT SUPPL_NO REC_NO
NEXT SUPPL_NAME REC_NO
1
has
N
LINE
N
used
for
each
1
PARTS
ORDER_NO
LINE_NO (13)
QTY
PART_NO
NEXT ORDER_NO REC_NO
NEXT LINE_NO REC_NO
PART_NO (50)
PART_NAME (8)
EMPLO_NO
SUPPL1_NO
SUPPL2_NO
NEXT PART_NO REC_NO
NEXT PART_NAME REC_NO

The Pre-Test:

1. Find and list all customers, by name [CUST_NAME], whose names begin with the letter "D".

Answer (s): ____________ ____________ ____________

TIME ===> ____________

2. Who was the customer, by name, for order no. 30?

ANSWER: ____________

TIME ===> ____________

3. Find and list the order numbers [NOTE: There are only THREE] for the two customers whose names begin with the letter "C".

ANSWERS: ____________ ____________ ____________

TIME ===> ____________

The Queries:

1. Find and list the customer, by name whose name begins with the letter "G".

ANSWER: ____________

TIME:===> ____________

2. Which two suppliers by name and number, handle the parts BUTTONS?

ANSWERS: ____________ ____________

TIME:===> ____________

3. Which part numbers were ordered on Order No. 105?

ANSWERS: ____________ ____________ ____________

____________ ____________ ____________

TIME:===> ____________

4. Employee No. 354 works with which suppliers (by name)?

ANSWERS: ____________ ____________ ____________

____________ ____________ ____________

TIME:===> ____________

5. What parts (by name) were ordered on Order No. 42?

ANSWERS: ____________ ____________ ____________

____________ ____________ ____________

TIME:===> ____________

6. What is the address of the customer who ordered on the date 07 February?

ANSWER: ____________

TIME:===> ____________

7. What are the two suppliers (by name) of the parts ordered on Order No. 112 ?

ANSWERS: _______________ _______________

TIME:===> _______________

8. What part (by name) was ordered in the month of March and who ordered it (by name)?

ANSWERS: _______________ _______________

TIME===> _______________

9. Is every Supplier assigned to at least one employee?

[ ] yes [ ] no

if "no", which supplier is unassigned? _______________

TIME:===> _______________

10. What parts (by part no.) were ordered in the month of January?

ANSWERS: _______________ _______________ _______________

_______________ _______________ _______________

TIME:===> _______________

APPENDIX E

Raw Data

The following pages include the actual raw data elements, based on performance of the individual subjects. The enclosed data include:

Ranked Pre-test Scores 118

Demographic Data by Group 119

Cumulative Times by Group 120

Net Times by Group (Test 1) 121

Effectiveness by Group (Test 1) 122

Cumulative Times by Group (Test 2) 123

Net Times by Group (Test 2) 124

Effectiveness by Group (Test 2) 125

Coefficients of Determination Matrix 126

RANKED PRE-TEST SCORES

1	2	3	4
3.00	72	ELDR	A
3.00	122	CUST	B
3.00	126	GARV	A
3.00	150	DUVA	B
2.67	43	HUNT	A
2.67	73	MCRA	B
2.67	85	BUCK	A
2.33	70	RIZZ	B
2.33	70	MCCL	A
2.33	78	LINE	B
2.33	87	GRIE	A
2.33	117	RUSS	B
2.33	137	DUMO	A
2.00	55	HUOT	B
2.00	60	LAYC	A
2.00	62	BALL	B
2.00	83	DIEF	A
2.00	88	TRUC	B
2.00	109	SHEA	A
2.00	110	ENOS	B
2.00	121	LATT	A
2.00	123	PERR	B
2.00	125	SOBO	A
2.00	145	DIZI	B
2.00	168	SERI	A
2.00	178	LABR	B
1.00	18	GAMA	A
1.00	99	RILE	B
3.00		STOK	A < Pre-Test not timed
		FRAS	B < Pre-Test not taken

COLUMN DESCRIPTIONS:
1 No of correct query responses.
2 Time to complete queries through last correct query.
3 Name code
4 Group assignment

DEMOGRAPHIC DATA BY GROUP

Subjects x Demographic Data Question No.

************** Group A *******

	1	2	3	4	5
BUCK	M	21	2	1	2.73
DIEF	M	21	1	1	2.36
DUMO	F	22	1	1	2.60
ELDR	M	21	3	1	3.70
GAMA	M	20	2	2	2.50
GARV	M	22	2	2	2.00
GRIE	F	20	1	2	2.70
HUNT	F	20	2	1	2.85
LATT	M	22	2	2	2.00
LAYC	M	22	2	1	2.40
MCCL	M	21	2	1	3.11
SERI	M	23	1	1	2.36
SHEA	M	21	2	1	3.20
SOBO	F	21	2	1	2.67
STOK	M	21	1	1	3.55
mean		21.2	1.73	1.27	2.72

************** Group B *******

	1	2	3	4	5
BALL	M	20	1	2	2.50
CUST	M	24	2	2	2.60
DIZI	M	20	2	2	2.50
DUVA	M	22	2	2	2.40
ENOS	F	21	1	1	2.34
FRAS	M	23	2	2	2.53
HUOT	M	20	2	2	2.80
LABR	M	23	2	1	2.42
LINE	M	21	2	2	2.70
MCRA	F	21	2	2	2.93
PERR	F	21	2	1	2.30
RILE	F	22	1	2	2.20
RIZZ	M	22	2	1	2.40
RUSS	M	21	2	1	2.86
TRUC	M	22	2	1	2.75
mean		21.5	1.8	1.6	2.55

Demographic Data
by Question No.

1. Male or Female
2. Age
3. Database exposure.
4. Math(1) vs. Language(2)
5. G.P.A.

CUMULATIVE TIMES BY GROUP (Test 1)

Subjects x Query No.

************** Group A ******************

	1	2	3	4	5	6	7	8	9	10
BUCK	7	54	169	185	267	314	385	440	457	509
DIEF	2	22	108	127	171	214	251	373	375	492
DUMO	24	52	113	132	268	330	466	494	500	600
ELDR	9	30	80	97	135	151	203	255	300	361
GAMA	58	99	110	167	230	252	421	530	545	595
GARV	11	34	137	224	331	354	416	480	519	583
GRIE	10	37	132	153	210	256	322	418	459	513
HUNT	11	38	82	98	141	156	206	265	364	490
LATT	13	94	160	174	312	354	430	540	588	600
LAYC	10	45	123	175	224	310	369	427	442	483
MCCL	15	31	95	132	173	188	257	337	371	503
SERI	14	84	210	260	400	490	600			
SHEA	10	53	109	213	259	303	384	425	503	565
SOBO	8	52	160	190	299	321	359	402	490	564
STOK	3	32	90	106	120	174	219	274	306	359

************** Group B ******************

	1	2	3	4	5	6	7	8	9	10
BALL	10	49	75	98	173	205	308	414	436	518
CUST	14	48	169	192	253	284	308	369	410	471
DIZI	50	79	130	154	172	214	423	500	543	600
DUVA	15	69	118	145	172	216	291	368	386	471
ENOS	10	32	111	136	200	217	288	382	442	508
FRAS	30	150	257	283	303	326	372	504	558	600
HUOT	10	38	74	97	163	204	308	413	436	517
LABR	14	46	73	96	188	194	232	294	323	370
LINE	11	33	180	220	265	311	336	498	514	540
MCRA	15	110	160	178	253	310	355	498	539	600
PERR	15	43	189	222	342	376	424	483	499	567
RILE	12	56	94	118	154	177	211	291	433	491
RIZZ	30	107	170	200	241	282	319	427	473	526
RUSS	3	24	118	201	339	404	459	600		
TRUC	11	32	106	135	169	188	231	364	399	469

NET TIMES BY GROUP (Test 1)

Subjects x Query No.

************** Group A ********************

	1	2	3	4	5	6	7	8	9	10
BUCK	7	47	115	16	82	47	71	55	17	52
DIEF	2	20	86	19	44	43	37	122	2	117
DUMO	24	28	61	19	136	62	136	28	6	100
ELDR	9	21	50	17	38	16	52	52	45	61
GAMA	58	41	11	57	63	22	169	109	15	50
GARV	11	23	103	87	107	23	62	64	39	64
GRIE	10	27	95	21	57	46	66	96	41	54
HUNT	11	27	44	16	43	15	50	59	99	126
LATT	13	81	66	14	138	42	76	110	48	12
LAYC	10	35	78	52	49	86	59	58	15	41
MCCL	15	16	64	37	41	15	69	80	34	132
SERI	14	70	126	50	140	90	110			
SHEA	10	43	56	104	46	44	81	41	78	62
SOBO	8	44	108	30	109	22	38	43	88	74
STOK	3	29	58	16	14	54	45	55	32	53
mean	14	37	75	37	74	42	75	69	40	71

mean of means 53.3

************** Group B ********************

	1	2	3	4	5	6	7	8	9	10
BALL	10	39	26	23	75	32	103	106	22	82
CUST	14	34	121	23	61	31	24	61	41	61
DIZI	50	29	51	24	18	42	209	77	43	57
DUVA	15	54	49	27	27	44	75	77	18	85
ENOS	10	22	79	25	64	17	71	94	60	66
FRAS	30	120	107	26	20	23	46	132	54	42
HUOT	10	28	36	23	66	41	104	105	23	81
LABR	14	32	27	23	92	6	38	62	29	47
LINE	11	22	147	40	45	46	25	162	16	26
MCRA	15	95	50	18	75	57	45	143	41	61
PERR	15	28	146	33	120	34	48	59	16	68
RILE	12	44	38	24	36	23	34	80	142	58
RIZZ	30	77	63	30	41	41	37	108	46	53
RUSS	3	21	94	83	138	65	55	141		
TRUC	11	21	74	29	34	19	43	133	35	70
mean	17	44	74	30	61	35	64	103	42	61

mean of means 53.0

EFFECTIVENESS BY GROUP (Test 1)

Subjects x Query No.

Data elements represent portion of query correctly retrieved.

************** Group A **************************

	1	2	3	4	5	6	7	8	9	10	mean
BUCK	1	1	1	1	1	1	1	.5	1	.5	90
DIEF	1	.7	.25	.67	.25	1	1	1	1	0	68.7
DUMO	1	.5	1	1	.75	1	.5	1	1	0	77.5
ELDR	1	1	1	1	.5	.5	1	1	1	1	90
GAMA	1	.3	.25	.67	.25	1	0	1	1	.4	58.2
GARV	1	1	.25	1	1	1	1	1	1	1	92.5
GRIE	1	0	.25	1	.25	1	1	1	1	.4	69
HUNT	1	1	.5	1	1	1	0	1	1	0	75
LATT	1	.5	.25	1	1	1	1	1	.5	0	72.5
LAYC	1	1	.25	1	.25	1	1	.5	1	.5	75
MCCL	1	1	1	1	1	1	1	1	1	1	100
SERI	1	1	1	.67	.5	1	0	0	0	0	51.7
SHEA	1	1	0	1	.25	1	1	1	0	0	62.5
SOBO	1	1	.25	1	.25	1	1	1	1	.4	79
STOK	1	1	1	1	1	1	1	.5	1	1	95
mean	1	.8	.55	.93	.62	.97	.77	.83	.83	.41	77.1
							std deviation				13.7

************** Group B **************************

	1	2	3	4	5	6	7	8	9	10	mean
BALL	1	1	0	1	.25	1	1	0	1	.5	67.5
CUST	1	1	.75	1	1	1	1	1	1	.7	94.5
DIZI	.5	1	1	1	1	1	.5	1	1	.4	84
DUVA	1	1	.25	1	.25	1	1	1	1	.5	80
ENOS	1	1	1	1	.5	1	1	.5	1	.5	85
FRAS	1	1	.25	1	.25	1	1	1	1	.3	78
HUOT	1	1	0	1	.25	1	1	.5	1	.5	72.5
LABR	1	1	.25	1	.25	0	1	.5	1	.5	65
LINE	1	1	.75	1	0	1	1	1	1	.2	79.5
MCRA	1	1	1	1	.25	.5	1	1	1	.5	82.5
PERR	1	0	.5	1	1	.5	1	1	1	.2	72
RILE	0	0	0	.67	.25	1	1	1	0	.5	44.2
RIZZ	1	1	.25	1	.25	1	1	1	1	.3	78
RUSS	1	1	1	1	.5	1	1	0	0	0	65
TRUC	1	.5	1	1	1	1	1	.5	1	1	90
mean	.9	.83	.53	.98	.47	.87	.97	.73	.87	.44	75.8
							std deviation				11.9

CUMULATIVE TIMES BY GROUP (Test 2)

Subjects x Query No.

************** Group A ******************

	1	2	3	4	5	6	7	8	9	10
BUCK	4	31	84	108	142	161	177	231	249	315
DIEF	9	27	48	70	100	117	177	290	292	333
DUMO	39	93	155	175	240	264	292	333	335	408
ELDR	9	50	91	115	153	168	189	210	226	286
GAMA	25	83	101	127	170	195	290	305	325	460
GARV	15	38	82	104	139	189	238	300	340	394
GRIE	8	92	134	152	173	191	224	291	294	355
HUNT	10	25	64	91	122	135	153	200	224	277
LATT	14	36	231	357	414	432	480	550	557	600
LAYC	10	30	82	98	141	166	200	284	296	340
MCCL	14	32	53	77	104	119	164	277	280	364
SERI	20	112	280	340	428	478	520	600		
SHEA	13	51	103	133	153	195	245	291	335	427
SOBO	12	39	66	83	101	117	179	231	296	351
STOK	11	28	75	89	108	124	149	222	230	298

************** Group B ******************

	1	2	3	4	5	6	7	8	9	10
BALL	10	37	89	110	156	200	231	311	339	395
CUST	8	26	69	99	125	137	162	207	224	297
DIZI	6	72	104	134	152	192	283	311	333	419
DUVA	24	42	67	87	104	123	150	220	234	330
ENOS	11	30	72	87	106	109	134	195	260	315
FRAS	12	34	50	70	83	100	131	238	336	570
HUOT	16	39	70	90	105	119	146	301	321	493
LABR	15	65	103	120	129	142	166	225	234	292
LINE	12	41	81	100	130	150	230	290	295	394
MCRA	15	33	55	72	86	127	146	199	209	248
PERR	15	33	103	124	138	173	195	223	253	301
RILE	21	59	156	175	223	238	272	331	350	396
RIZZ	20	46	88	115	147	177	207	303	320	367
RUSS	8	34	60	103	125	140	158	197	216	277
TRUC	9	25	43	58	79	93	129	171	213	295

NET TIMES BY GROUP (Test 2)

Subjects x Query No.

************** Group A ******************

	1	2	3	4	5	6	7	8	9	10
BUCK	4	27	53	24	34	19	16	54	18	66
DIEF	9	18	21	22	30	17	60	113	2	41
DUMO	39	54	62	20	65	24	28	41	2	73
ELDR	9	41	41	24	38	15	21	21	16	60
GAMA	25	58	18	26	43	25	95	15	20	135
GARV	15	23	44	22	35	50	49	62	40	54
GRIE	8	84	42	18	21	18	33	67	3	61
HUNT	10	15	39	27	31	13	18	47	24	53
LATT	14	22	195	126	57	18	48	70	7	43
LAYC	10	20	52	16	43	25	34	84	12	44
MCCL	14	18	21	24	27	15	45	113	3	84
SERI	20	92	168	60	88	50	42	80		
SHEA	13	38	52	30	20	42	50	46	44	92
SOBO	12	27	27	17	18	16	62	52	65	55
STOK	11	17	47	14	19	16	25	73	8	68
mean	14	37	59	31	38	24	42	63	19	66

mean of means 39.2
std deviation 30.7

************** Group B ******************

	1	2	3	4	5	6	7	8	9	10
BALL	10	27	52	21	46	44	31	80	28	56
CUST	8	18	43	30	26	12	25	45	17	73
DIZI	6	66	32	30	18	40	91	28	22	86
DUVA	24	18	25	20	17	19	27	70	14	96
ENOS	11	19	42	15	19	3	25	61	65	55
FRAS	12	22	16	20	13	17	31	107	98	234
HUOT	16	23	31	20	15	14	27	155	20	172
LABR	15	50	38	17	9	13	24	59	9	58
LINE	12	29	40	19	30	20	80	60	5	99
MCRA	15	18	22	17	14	41	19	53	10	39
PERR	15	18	70	21	14	35	22	28	30	48
RILE	21	38	97	19	48	15	34	59	19	46
RIZZ	20	26	42	27	32	30	30	96	17	47
RUSS	8	26	26	43	22	15	18	39	19	61
TRUC	9	16	18	15	21	14	36	42	42	82
mean	13	28	40	22	23	22	35	65	28	83

mean of means 36
std deviation 31.6

EFFECTIVENESS BY GROUP (Test 2)

Subjects x Query No.

Data elements represent portion of query correctly retrieved

************** Group A ****************************

	1	2	3	4	5	6	7	8	9	10	
BUCK	1	1	1	1	1	1	1	1	1	1	100
DIEF	1	1	1	1	1	1	1	1	1	.5	95
DUMO	1	1	1	1	1	1	1	.5	1	.3	88
ELDR	1	1	0	1	1	1	1	1	1	.8	88
GAMA	1	1	.25	.67	.75	1	.5	.5	1	.2	68.7
GARV	1	1	1	1	1	1	1	1	1	.8	98
GRIE	1	1	1	1	1	1	1	1	1	.5	95
HUNT	1	1	1	1	1	1	1	1	1	.5	95
LATT	0	1	1	1	1	1	1	1	0	0	70
LAYC	1	1	.25	1	.25	1	0	1	0	.5	60
MCCL	1	1	.25	.67	1	1	1	1	1	.8	87.2
SERI	1	1	1	1	1	1	0	0	0	0	60
SHEA	1	1	.25	1	0	1	1	.5	0	.2	59.5
SOBO	1	1	.25	1	.25	1	1	1	1	.5	80
STOK	1	1	1	1	1	1	1	1	1	1	100
mean	.93	1.0	.68	.96	.82	1.0	.83	.83	.73	.51	82.96
								std dev =			15

************** Group B ****************************

	1	2	3	4	5	6	7	8	9	10	
BALL	1	1	.25	1	.5	0	.5	1	1	.4	66.5
CUST	1	1	1	1	1	1	1	1	1	1	100
DIZI	1	1	.25	.67	.75	1	0	.5	1	0	61.7
DUVA	1	1	.25	1	.25	0	1	1	1	.6	71
ENOS	1	1	1	1	1	1	1	.5	1	.4	89
FRAS	1	1	.25	1	.25	1	1	1	1	.4	79
HUOT	1	1	.25	1	.25	1	1	1	0	.5	70
LABR	1	1	.25	1	.25	1	1	1	1	.5	80
LINE	1	1	.25	.67	.25	1	1	1	1	0	71.7
MCRA	1	1	.25	1	.25	1	1	1	1	.5	80
PERR	1	1	.5	1	1	1	1	1	1	.5	90
RILE	1	1	1	1	1	1	1	.5	0	0	75
RIZZ	1	1	.25	1	.25	1	1	1	1	0	75
RUSS	1	0	.25	1	.25	1	1	1	1	.5	70
TRUC	1	1	1	1	1	1	1	1	1	1	100
mean	1.0	.93	.47	.96	.55	.87	.90	.90	.87	.42	78.59
									std dev=		11

COEFFICIENTS OF DETERMINATION MATRIX
(r-squares)

Demographic Factors x Test Measures of Efficiency and Effectiveness

	Test 1 Effic'y	Test 2 Effect	Test 1 Effic'y	Test 2 Effect
Gender	.00129	.02638	.06947	.06733
Age	.00296	.00151	.04400	.00532
Database Exper.	.02386	.12270	.03940	.00580
Cogn. Affinity	.00838	.01026	.07414	.04839
GPA	.11750	.13138	.12306	.03389

APPENDIX F

The Database

The following database is provided in its entirety in order to illustrate the scope and relative complexity of the experimental database. The data elements are identical to those used in the test. The format is similar to that used on the 3 x 5 cards in the experiment which had one record per card.

Further discussion of the nature of this database is found in Chapter 4, Section C.

```
001              ORDER                        001

    ORDER_NO:    545
    CUST_NO:     036
    DATE:        03 SEP 88

    NEXT ORDER_NO REC_NO:   123
    THIS CUST_NO REC_NO:    031
    NEXT DATE REC_NO:       095
```

```
002              ORDER                        002

    ORDER_NO:    811
    CUST_NO:     134
    DATE:        08 JUN 88

    NEXT ORDER_NO REC_NO:   064
    THIS CUST_NO REC_NO:    130
    NEXT DATE REC_NO:       054
```

```
003              PARTS                        003

    PART_NO:     343
    PART_NAME:   WASHER
    EMPLO_NO:    127 REC_NO: 040

    SUPPL1_NO:   087 REC_NO: 065
    SUPPL2_NO:   065 REC_NO: 057
    NEXT PART_NO REC_NO:     073
    NEXT PART NAME REC_NO:   052
```

```
004              ORDER                        004

    ORDER_NO:    216
    CUST_NO:     946
    DATE:        14 APR 88

    NEXT ORDER_NO REC_NO:   103
    THIS CUST_NO REC_NO:    114
    NEXT DATE REC_NO:       044
```

```
005             LINE                      005

    ORDER_NO:    543
    LINE_NO:     001
    QTY:         4

    PART_NO:     933 REC_NO: 008
    NEXT LINE_NO REC_NO:     061
    NEXT ORDER_NO REC_NO:    087
```

```
006             LINE                      006

    ORDER_NO:    332
    LINE_NO:     001
    QTY:         2

    PART_NO:     157 REC_NO: 043
    NEXT LINE_NO REC_NO:     113
    NEXT ORDER_NO REC_NO:    042
```

```
007             PARTS                     007

    PART_NO:     116
    PART_NAME:   PIN
    EMPLO_NO:    443 REC_NO: 067

    SUPPL1_NO:   021 REC_NO: 062
    SUPPL2_NO:   043 REC_NO: 104
    NEXT PART_NO REC_NO:     019
    NEXT PART NAME REC_NO:   129
```

```
008             PARTS                     008

    PART_NO:     933
    PART_NAME:   ARM
    EMPLO_NO:    238 REC_NO: 099

    SUPPL1_NO:   087 REC_NO: 065
    SUPPL2_NO:   043 REC_NO: 104
    NEXT PART_NO REC_NO:     EOF
    NEXT PART NAME REC_NO:   019
```

```
009           LINE                              009

    ORDER_NO:    535
    LINE_NO:     001
    QTY:         4

    PART_NO:     312 REC_NO: 046
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    005
```

```
010           ORDER                             010

    ORDER_NO:    323
    CUST_NO:     878
    DATE:        16 OCT 88

    NEXT ORDER_NO REC_NO:    118
    THIS CUST_NO REC_NO:     091
    NEXT DATE REC_NO:        041
```

```
011           CUST                              011

    CUST_NO:     612
    CUST_NAME:   ABRIAL
    ADDRESS:     BOSTON

    NEXT CUST_NO REC_NO:     105
    NEXT CUST_NAME REC_NO:   114
```

```
012           LINE                              012

    ORDER_NO:    647
    LINE_NO:     001
    QTY:         4

    PART_NO:     718 REC_NO: 132
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    092
```

```
013            LINE                     013

    ORDER_NO:     010
    LINE_NO:      001
    QTY:          2

    PART_NO:      116 REC_NO: 007
    NEXT LINE_NO REC_NO:      EOF
    NEXT ORDER_NO REC_NO:     131
```

```
014            LINE                     014

    ORDER_NO:     042
    LINE_NO:      004
    QTY:          2

    PART_NO:      053 REC_NO: 033
    NEXT LINE_NO REC_NO:      EOF
    NEXT ORDER_NO REC_NO:     032
```

```
015            CUST                     015

    CUST_NO:      274
    CUST_NAME:    GIBBS
    ADDRESS:      SANDWICH

    NEXT CUST_NO REC_NO:      117
    NEXT CUST_NAME REC_NO:    090
```

```
016            CUST                     016

    CUST_NO:      922
    CUST_NAME:    CODD
    ADDRESS:      NORTHFIELD

    NEXT CUST_NO REC_NO:      114
    NEXT CUST_NAME REC_NO:    110
```

```
017              PARTS                        017

    PART_NO:      041
    PART_NAME:    KNOB
    EMPLO_NO:     354 REC_NO: 025

    SUPPL1_NO:    065 REC_NO: 057
    SUPPL2_NO:    010 REC_NO: 049
    NEXT PART_NO REC_NO:      033
    NEXT PART NAME REC_NO:    050
```

```
018              LINE                         018

    ORDER_NO:     146
    LINE_NO:      001
    QTY:          3

    PART_NO:      451 REC_NO: 119
    NEXT LINE_NO REC_NO:      039
    NEXT ORDER_NO REC_NO:     035
```

```
019              PARTS                        019

    PART_NO:      123
    PART_NAME:    BELT
    EMPLO_NO:     443 REC_NO: 067

    SUPPL1_NO:    021 REC_NO: 062
    SUPPL2_NO:    087 REC_NO: 065
    NEXT PART_NO REC_NO:      043
    NEXT PART NAME REC_NO:    133
```

```
020              LINE                         020

    ORDER_NO:     908
    LINE_NO:      001
    QTY:          3

    PART_NO:      343 REC_NO: 003
    NEXT LINE_NO REC_NO:      EOF
    NEXT ORDER_NO REC_NO:     128
```

```
021            LINE                      021

    ORDER_NO:    229
    LINE_NO:     001
    QTY:         1

    PART_NO:     116 REC_NO: 007
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    071
```

```
022            LINE                      022

    ORDER_NO:    216
    LINE_NO:     002
    QTY:         3

    PART_NO:     041 REC_NO: 017
    NEXT LINE_NO REC_NO:     116
    NEXT ORDER_NO REC_NO:    038
```

```
023            LINE                      023

    ORDER_NO:    216
    LINE_NO:     004
    QTY:         4

    PART_NO:     658 REC_NO: 129
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    038
```

```
024            CUST                      024

    CUST_NO:     556
    CUST_NAME:   BRODIE
    ADDRESS:     PLYMOUTH

    NEXT CUST_NO REC_NO:     094
    NEXT CUST_NAME REC_NO:   031
```

```
025             EMPLOYEE                          025

    EMPLO_NO:    354
    SUPPL1_NO:   021 REC_NO: 062
    SUPPL2_NO:   065 REC_NO: 057

    SUPPL3_NO:   010 REC_NO: 049


    NEXT EMPLO_NO REC_NO:   067
```

```
026             LINE                              026

    ORDER_NO:    624
    LINE_NO:     001
    QTY:         1

    PART_NO:     041 REC_NO: 017
    NEXT LINE_NO REC_NO:    EOF
    NEXT ORDER_NO REC_NO:   012
```

```
027             CUST                              027

    CUST_NO:     122
    CUST_NAME:   ZLOOF
    ADDRESS:     BRISTOL

    NEXT CUST_NO REC_NO:    130
    NEXT CUST_NAME REC_NO:  EOF
```

```
028             LINE                              028

    ORDER_NO:    738
    LINE_NO:     001
    QTY:         1

    PART_NO:     157 REC_NO: 043
    NEXT LINE_NO REC_NO:    EOF
    NEXT ORDER_NO REC_NO:   077
```

```
029            LINE                     029

    ORDER_NO:    738
    LINE_NO:     001
    QTY:         1

    PART_NO:     237 REC_NO: 074
    NEXT LINE_NO REC_NO:     028
    NEXT ORDER_NO REC_NO:    077
```

```
030            ORDER                    030

    ORDER_NO:    146
    CUST_NO:     564
    DATE:        12 AUG 88

    NEXT ORDER_NO REC_NO:    004
    THIS CUST_NO REC_NO:     094
    NEXT DATE REC_NO:        076
```

```
031            CUST                     031

    CUST_NO:     036
    CUST_NAME:   CHEN
    ADDRESS:     NORFOLK

    NEXT CUST_NO REC_NO:     027
    NEXT CUST_NAME REC_NO:   016
```

```
032            LINE                     032

    ORDER_NO:    105
    LINE_NO:     001
    QTY:         1

    PART_NO:     451 REC_NO: 119
    NEXT LINE_NO REC_NO:     060
    NEXT ORDER_NO REC_NO:    072
```

```
033              PARTS                          033

    PART_NO:     053
    PART_NAME:   BOLT
    EMPLO_NO:    354 REC_NO: 025

    SUPPL1_NO:   021 REC_NO: 062
    SUPPL2_NO:   065 REC_NO: 057
    NEXT PART_NO REC_NO:     007
    NEXT PART NAME REC_NO:   126
```

```
034              LINE                           034

    ORDER_NO:    421
    LINE_NO:     003
    QTY:         3

    PART_NO:     933 REC_NO: 008
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    053
```

```
035              LINE                           035

    ORDER_NO:    216
    LINE_NO:     001
    QTY:         2

    PART_NO:     237 REC_NO: 074
    NEXT LINE_NO REC_NO:     022
    NEXT ORDER_NO REC_NO:    038
```

```
036              LINE                           036

    ORDER_NO:    836
    LINE_NO:     001
    QTY:         3

    PART_NO:     546 REC_NO: 126
    NEXT LINE_NO REC_NO:     097
    NEXT ORDER_NO REC_NO:    020
```

```
037            LINE                     037

    ORDER_NO:    440
    LINE_NO:     002
    QTY:         3

    PART_NO:     432 REC_NO: 073
    NEXT LINE_NO REC_NO:     055
    NEXT ORDER_NO REC_NO:    009
```

```
038            LINE                     038

    ORDER_NO:    226
    LINE_NO:     001
    QTY:         2

    PART_NO:     556 REC_NO: 106
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    021
```

```
039            LINE                     039

    ORDER_NO:    146
    LINE_NO:     002
    QTY:         3

    PART_NO:     312 REC_NO: 046
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    035
```

```
040            EMPLOYEE                 040

    EMPLO_NO:    127
    SUPPL1_NO:   087 REC_NO: 065
    SUPPL2_NO:   065 REC_NO: 057

    SUPPL3_NO:   021 REC_NO: 062
    SUPPL4_NO:   043 REC_NO: 104
    SUPPL5_NO:   010 REC_NO: 049
    NEXT EMPLO_NO REC_NO:    099
```

```
041              ORDER                              041

    ORDER_NO:     332
    CUST_NO:      028
    DATE:         21 OCT 88

    NEXT ORDER_NO REC_NO:    118
    THIS CUST_NO REC_NO:     108
    NEXT DATE REC_NO:        115
```

```
042              LINE                               042

    ORDER_NO:     421
    LINE_NO:      001
    QTY:          4

    PART_NO:      556 REC_NO: 106
    NEXT LINE_NO REC_NO:     056
    NEXT ORDER_NO REC_NO:    053
```

```
043              PARTS                              043

    PART_NO:      157
    PART_NAME:    SWITCH
    EMPLO_NO:     443 REC_NO: 067

    SUPPL1_NO:    043 REC_NO: 104
    SUPPL2_NO:    021 REC_NO: 062
    NEXT PART_NO REC_NO:     089
    NEXT PART NAME REC_NO:   003
```

```
044              ORDER                              044

    ORDER_NO:     707
    CUST_NO:      946
    DATE:         01 MAY 88

    NEXT ORDER_NO REC_NO:    081
    THIS CUST_NO REC_NO:     114
    NEXT DATE REC_NO:        120
```

```
045            ORDER                    045

    ORDER_NO:    535
    CUST_NO:     922
    DATE:        07 FEB 88

    NEXT ORDER_NO REC_NO:    076
    THIS CUST_NO REC_NO:     016
    NEXT DATE REC_NO:        078
```

```
046            PARTS                    046

    PART_NO:     312
    PART_NAME:   WIRE
    EMPLO_NO:    127 REC_NO: 040

    SUPPL1_NO:   043 REC_NO: 104
    SUPPL2_NO:   065 REC_NO: 057
    NEXT PART_NO REC_NO:     084
    NEXT PART NAME REC_NO:   EOF
```

```
047            CUST                     047

    CUST_NO:     806
    CUST_NAME:   MINSKY
    ADDRESS:     BEDFORD

    NEXT CUST_NO REC_NO:     091
    NEXT CUST_NAME REC_NO:   109
```

```
048            CUST                     048

    CUST_NO:     376
    CUST_NAME:   DEXTER
    ADDRESS:     YORK

    NEXT CUST_NO REC_NO:     110
    NEXT CUST_NAME REC_NO:   015
```

```
053            LINE                      053

    ORDER_NO:    439
    LINE_NO:     001
    QTY:         1

    PART_NO:     759 REC_NO: 121
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    100
```

```
054            ORDER                     054

    ORDER_NO:    624
    CUST_NO:     612
    DATE:        19 JUN 88

    NEXT ORDER_NO REC_NO:    078
    THIS CUST_NO REC_NO:     011
    NEXT DATE REC_NO:        081
```

```
055            LINE                      055

    ORDER_NO:    440
    LINE_NO:     003
    QTY:         4

    PART_NO:     343 REC_NO: 003
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    009
```

```
056            LINE                      056

    ORDER_NO:    421
    LINE_NO:     002
    QTY:         2

    PART_NO:     759 REC_NO: 121
    NEXT LINE_NO REC_NO:     034
    NEXT ORDER_NO REC_NO:    053
```

```
057             SUPPLIER                    057

    SUPPL_NO:    065
    SUPPL_NAME:  GOODE


    NEXT SUPPL_NO REC_NO :  065
    NEXT SUPPL_NAME REC_NO: 065
```

```
058             LINE                        058

    ORDER_NO:    827
    LINE_NO:     001
    QTY:         2

    PART_NO:     334 REC_NO: 084
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    036
```

```
059             LINE                        059

    ORDER_NO:    042
    LINE_NO:     002
    QTY:         3

    PART_NO:     432 REC_NO: 073
    NEXT LINE_NO REC_NO:     088
    NEXT ORDER_NO REC_NO:    032
```

```
060             LINE                        060

    ORDER_NO:    105
    LINE_NO:     002
    QTY:         1

    PART_NO:     240 REC_NO: 085
    NEXT LINE_NO REC_NO:     124
    NEXT ORDER_NO REC_NO:    072
```

```
061            LINE                        061

    ORDER_NO:   543
    LINE_NO:    002
    QTY:        1

    PART_NO:    822 REC_NO: 133
    NEXT LINE_NO REC_NO:    EOF
    NEXT ORDER_NO REC_NO:   087
```

```
062            SUPPLIER                    062

    SUPPL_NO:   021
    SUPPL_NAME: ACE


    NEXT SUPPL_NO REC_NO :  104
    NEXT SUPPL_NAME REC_NO: 104
```

```
063            LINE                        063

    ORDER_NO:   811
    LINE_NO:    001
    QTY:        1

    PART_NO:    021 REC_NO: 050
    NEXT LINE_NO REC_NO:    079
    NEXT ORDER_NO REC_NO:   058
```

```
064            ORDER                       064

    ORDER_NO:   827
    CUST_NO:    678
    DATE:       02 JUL 88

    NEXT ORDER_NO REC_NO:   127
    THIS CUST_NO REC_NO:    105
    NEXT DATE REC_NO:       123
```

```
065              SUPPLIER                          065

     SUPPL_NO:     087
     SUPPL_NAME:  FINE


     NEXT SUPPL_NO REC_NO :  EOF
     NEXT SUPPL_NAME REC_NO: 049
```

```
066              ORDER                             066

     ORDER_NO:     105
     CUST_NO:      478
     DATE:         10 DEC 88

     NEXT ORDER_NO REC_NO:   098
     THIS CUST_NO REC_NO:    090
     NEXT DATE REC_NO:       135
```

```
067              EMPLOYEE                          067

     EMPLO_NO:     443
     SUPPL1_NO:    021 REC_NO: 062
     SUPPL2_NO:    043 REC_NO: 104

     SUPPL3_NO:    087 REC_NO: 065
     SUPPL4_NO:    065 REC_NO: 057

     NEXT EMPLO_NO REC_NO:   EOF
```

```
068              LINE                              068

     ORDER_NO:     105
     LINE_NO:      004
     QTY:          1

     PART_NO:      847 REC_NO: 052
     NEXT LINE_NO REC_NO:    EOF
     NEXT ORDER_NO REC_NO:   072
```

```
069            LINE                      069

    ORDER_NO:    332
    LINE_NO:     003
    QTY:         3

    PART_NO:     450 REC_NO: 083
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    042
```

```
070            LINE                      070

    ORDER_NO:    614
    LINE_NO:     001
    QTY:         4

    PART_NO:     123 REC_NO: 019
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    026
```

```
071            LINE                      071

    ORDER_NO:    301
    LINE_NO:     001
    QTY:         1

    PART_NO:     227 REC_NO: 089
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    006
```

```
072            LINE                      072

    ORDER_NO:    112
    LINE_NO:     001
    QTY:         2

    PART_NO:     053 REC_NO: 033
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    018
```

---

073 PARTS 073

```
PART_NO:      432
PART_NAME:    SPACER
EMPLO_NO:     238 REC_NO: 099

SUPPL1_NO:    087 REC_NO: 065
SUPPL2_NO:    010 REC_NO: 049
NEXT PART_NO REC_NO:      083
NEXT PART NAME REC_NO:    119
```

---

074 PARTS 074

```
PART_NO:      237
PART_NAME:    HOOK
EMPLO_NO:     064 REC_NO: 093

SUPPL1_NO:    043 REC_NO: 104
SUPPL2_NO:    021 REC_NO: 062
NEXT PART_NO REC_NO:      085
NEXT PART NAME REC_NO:    017
```

---

075 PARTS 075

```
PART_NO:      919
PART_NAME:    GEAR
EMPLO_NO:     238 REC_NO: 099

SUPPL1_NO:    087 REC_NO: 065
SUPPL2_NO:    010 REC_NO: 049
NEXT PART_NO REC_NO:      008
NEXT PART NAME REC_NO:    074
```

---

076 ORDER 076

```
ORDER_NO:     543
CUST_NO:      206
DATE:         23 AUG 88

NEXT ORDER_NO REC_NO:     001
THIS CUST_NO REC_NO:      122
NEXT DATE REC_NO:         127
```

```
077             LINE                          077

    ORDER_NO:    748
    LINE_NO:     001
    QTY:         3

    PART_NO:     847 REC_NO: 052
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    063
```

```
078             ORDER                         078

    ORDER_NO:    647
    CUST_NO:     376
    DATE:        25 MAR 88

    NEXT ORDER_NO REC_NO:    044
    THIS CUST_NO REC_NO:     048
    NEXT DATE REC_NO:        004
```

```
079             LINE                          079

    ORDER_NO:    811
    LINE_NO:     002
    QTY:         2

    PART_NO:     919 REC_NO: 075
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    058
```

```
080             ORDER                         080

    ORDER_NO:    042
    CUST_NO:     806
    DATE:        16 JAN 88

    NEXT ORDER_NO REC_NO:    066
    THIS CUST_NO REC_NO:     047
    NEXT DATE REC_NO:        112
```

```
081            ORDER                      081

    ORDER_NO:     738
    CUST_NO:      739
    DATE:         20 JUN 88

    NEXT ORDER_NO REC_NO:    111
    THIS CUST_NO REC_NO:     096
    NEXT DATE REC_NO:        064
```

```
082            ORDER                      082

    ORDER_NO:     229
    CUST_NO:      206
    DATE:         13 MAY 88

    NEXT ORDER_NO REC_NO:    101
    THIS CUST_NO REC_NO:     122
    NEXT DATE REC_NO:        098
```

```
083            PARTS                      083

    PART_NO:      450
    PART_NAME:    BUTTON
    EMPLO_NO:     238 REC_NO: 099

    SUPPL1_NO:    087 REC_NO: 065
    SUPPL2_NO:    021 REC_NO: 062
    NEXT PART_NO REC_NO:      119
    NEXT PART NAME REC_NO:    132
```

```
084            PARTS                      084

    PART_NO:      334
    PART_NAME:    SNAP
    EMPLO_NO:     127 REC_NO: 040

    SUPPL1_NO:    065 REC_NO: 057
    SUPPL2_NO:    010 REC_NO: 049
    NEXT PART_NO REC_NO:      003
    NEXT PART NAME REC_NO:    073
```

```
085             PARTS                        085

    PART_NO:     240
    PART_NAME:   COVER
    EMPLO_NO:    064 REC_NO: 093

    SUPPL1_NO:   043 REC_NO: 104
    SUPPL2_NO:   065 REC_NO: 057
    NEXT PART_NO REC_NO:     046
    NEXT PART NAME REC_NO:   134
```

```
086             LINE                         086

    ORDER_NO:    042
    LINE_NO:     001
    QTY:         2

    PART_NO:     625 REC_NO: 134
    NEXT LINE_NO REC_NO:     059
    NEXT ORDER_NO REC_NO:    032
```

```
087             LINE                         087

    ORDER_NO:    545
    LINE_NO:     001
    QTY:         1

    PART_NO:     625 REC_NO: 134
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    070
```

```
088             LINE                         088

    ORDER_NO:    042
    LINE_NO:     003
    QTY:         1

    PART_NO:     227 REC_NO: 089
    NEXT LINE_NO REC_NO:     014
    NEXT ORDER_NO REC_NO:    032
```

```
089              PARTS                         089

    PART_NO:      227
    PART_NAME:    NUT
    EMPLO_NO:     064  REC_NO:  093

    SUPPL1_NO:    010  REC_NO:  049
    SUPPL2_NO:    043  REC_NO:  104
    NEXT PART_NO REC_NO:        074
    NEXT PART NAME REC_NO:      007
```

```
090              CUST                          090

    CUST_NO:      478
    CUST_NAME:    HAMMER
    ADDRESS:      CAMBRIDGE

    NEXT CUST_NO REC_NO:        024
    NEXT CUST_NAME REC_NO:      091
```

```
091              CUST                          091

    CUST_NO:      878
    CUST_NAME:    KENT
    ADDRESS:      CONCORD

    NEXT CUST_NO REC_NO:        016
    NEXT CUST_NAME REC_NO:      047
```

```
092              LINE                          092

    ORDER_NO:     707
    LINE_NO:      001
    QTY:          4

    PART_NO:      919  REC_NO:  075
    NEXT LINE_NO REC_NO:        EOF
    NEXT ORDER_NO REC_NO:       029
```

```
093             EMPLOYEE                        093

    EMPLO_NO:    064
    SUPPL1_NO:   010 REC_NO: 049
    SUPPL2_NO:   043 REC_NO: 104

    SUPPL3_NO:   065 REC_NO: 057
    SUPPL4_NO:   021 REC_NO: 062

    NEXT EMPLO_NO REC_NO:   040
```

```
094             CUST                            094

    CUST_NO:     564
    CUST_NAME:   SMITH
    ADDRESS:     CAMPTON

    NEXT CUST_NO REC_NO:     011
    NEXT CUST_NAME REC_NO:   117
```

```
095             ORDER                           095

    ORDER_NO:    941
    CUST_NO:     404
    DATE:        04 SEP 88

    NEXT ORDER_NO REC_NO:   EOF
    THIS CUST_NO REC_NO:    110
    NEXT DATE REC_NO:       107
```

```
096             CUST                            096

    CUST_NO:     739
    CUST_NAME:   NOLAN
    ADDRESS:     AUCKLAND

    NEXT CUST_NO REC_NO:     047
    NEXT CUST_NAME REC_NO:   122
```

```
097            LINE                        097

    ORDER_NO:    836
    LINE_NO:     002
    QTY:         3

    PART_NO:     718 REC_NO: 132
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    020
```

```
098            ORDER                       098

    ORDER_NO:    112
    CUST_NO:     122
    DATE:        17 MAY 88

    NEXT ORDER_NO REC_NO:    030
    THIS CUST_NO REC_NO:     027
    NEXT DATE REC_NO:        111
```

```
099            EMPLOYEE                    099

    EMPLO_NO:    238
    SUPPL1_NO:   087 REC_NO: 065
    SUPPL2_NO:   010 REC_NO: 049

    SUPPL3_NO:   043 REC_NO: 104
    SUPPL4_NO:   021 REC_NO: 062

    NEXT EMPLO_NO REC_NO:    025
```

```
100            LINE                        100

    ORDER_NO:    440
    LINE_NO:     001
    QTY:         2

    PART_NO:     546 REC_NO: 126
    NEXT LINE_NO REC_NO:     037
    NEXT ORDER_NO REC_NO:    009
```

```
101             ORDER                    101

    ORDER_NO:    301
    CUST_NO:     612
    DATE:        26 JUL 88

    NEXT ORDER_NO REC_NO:    010
    THIS CUST_NO REC_NO:     011
    NEXT DATE REC_NO:        030
```

```
102             ORDER                    102

    ORDER_NO:    440
    CUST_NO:     556
    DATE:        31 JAN 88

    NEXT ORDER_NO REC_NO:    045
    THIS CUST_NO REC_NO:     024
    NEXT DATE REC_NO:        045
```

```
103             ORDER                    103

    ORDER_NO:    226
    CUST_NO:     556
    DATE:        27 JAN 88

    NEXT ORDER_NO REC_NO:    082
    THIS CUST_NO REC_NO:     024
    NEXT DATE REC_NO:        102
```

```
104             SUPPLIER                 104

    SUPPL_NO:   043
    SUPPL_NAME: BEST


    NEXT SUPPL_NO REC_NO :  057
    NEXT SUPPL_NAME REC_NO: 057
```

```
105             CUST                          105

    CUST_NO:      678
    CUST_NAME:    SWENSON
    ADDRESS:      LONDON

    NEXT CUST_NO REC_NO:      109
    NEXT CUST_NAME REC_NO:    108
```

```
106             PARTS                         106

    PART_NO:      556
    PART_NAME:    COIL
    EMPLO_NO:     354 REC_NO: 025

    SUPPL1_NO:    010 REC_NO: 049
    SUPPL2_NO:    021 REC_NO: 062
    NEXT PART_NO REC_NO:      134
    NEXT PART NAME REC_NO:    085
```

```
107             ORDER                         107

    ORDER_NO:     030
    CUST_NO:      612
    DATE:         30 SEP 88

    NEXT ORDER_NO REC_NO:     080
    THIS CUST_NO REC_NO:      011
    NEXT DATE REC_NO:         010
```

```
108             CUST                          108

    CUST_NO:      028
    CUST_NAME:    TAYLOR
    ADDRESS:      HARTFORD

    NEXT CUST_NO REC_NO:      031
    NEXT CUST_NAME REC_NO:    130
```

```
109             CUST                        109

    CUST_NO:      709
    CUST_NAME:    NEWELL
    ADDRESS:      SPRINGFIELD

    NEXT CUST_NO REC_NO:      096
    NEXT CUST_NAME REC_NO:    096
```

```
110             CUST                        110

    CUST_NO:      404
    CUST_NAME:    DATE
    ADDRESS:      WELLINGTON

    NEXT CUST_NO REC_NO:      090
    NEXT CUST_NAME REC_NO:    048
```

```
111             ORDER                       111

    ORDER_NO:     748
    CUST_NO:      028
    DATE:         05 JUN 88

    NEXT ORDER_NO REC_NO:     002
    THIS CUST_NO REC_NO:      108
    NEXT DATE REC_NO:         002
```

```
112             ORDER                       112

    ORDER_NO:     439
    CUST_NO:      036
    DATE:         24 JAN 88

    NEXT ORDER_NO REC_NO:     102
    THIS CUST_NO REC_NO:      031
    NEXT DATE REC_NO:         103
```

113	LINE	113

```
ORDER_NO:    332
LINE_NO:     002
QTY:         1

PART_NO:     334 REC_NO: 084
NEXT LINE_NO REC_NO:     069
NEXT ORDER_NO REC_NO:    042
```

114	CUST	114

```
CUST_NO:     946
CUST_NAME:   BELKIN
ADDRESS:     MANCHESTER

NEXT CUST_NO REC_NO:     EOF
NEXT CUST_NAME REC_NO:   024
```

115	ORDER	115

```
ORDER_NO:    922
CUST_NO:     352
DATE:        28 NOV 88

NEXT ORDER_NO REC_NO:    095
THIS CUST_NO REC_NO:     117
NEXT DATE REC_NO:        118
```

116	LINE	116

```
ORDER_NO:    216
LINE_NO:     003
QTY:         2

PART_NO:     822 REC_NO: 133
NEXT LINE_NO REC_NO:     023
NEXT ORDER_NO REC_NO:    038
```

```
117             CUST                       117

    CUST_NO:     352
    CUST_NAME:   SOWA
    ADDRESS:     DOVER

    NEXT CUST_NO REC_NO:     048
    NEXT CUST_NAME REC_NO:   105
```

```
118             ORDER                      118

    ORDER_NO:    421
    CUST_NO:     709
    DATE:        06 DEC 88

    NEXT ORDER_NO REC_NO:    112
    THIS CUST_NO REC_NO:     109
    NEXT DATE REC_NO:        066
```

```
119             PARTS                      119

    PART_NO:     451
    PART_NAME:   SPRING
    EMPLO_NO:    238 REC_NO: 099

    SUPPL1_NO:   043 REC_NO: 104
    SUPPL2_NO:   087 REC_NO: 065
    NEXT PART_NO REC_NO:     126
    NEXT PART NAME REC_NO:   043
```

```
120             ORDER                      120

    ORDER_NO:    908
    CUST_NO:     709
    DATE:        19 MAY 88

    NEXT ORDER_NO REC_NO:    115
    THIS CUST_NO REC_NO:     109
    NEXT DATE REC_NO:        082
```

```
121             PARTS                    121

    PART_NO:      759
    PART_NAME:    SCREW
    EMPLO_NO:     064 REC_NO: 093

    SUPPL1_NO:    010 REC_NO: 049
    SUPPL2_NO:    043 REC_NO: 104
    NEXT PART_NO REC_NO:      133
    NEXT PART NAME REC_NO:    084
```

```
122             CUST                     122

    CUST_NO:      206
    CUST_NAME:    SHIPMAN
    ADDRESS:      WORCESTER

    NEXT CUST_NO REC_NO:      015
    NEXT CUST_NAME REC_NO:    094
```

```
123             ORDER                    123

    ORDER_NO:     614
    CUST_NO:      274
    DATE:         22 JUL 88

    NEXT ORDER_NO REC_NO:     054
    THIS CUST_NO REC_NO:      015
    NEXT DATE REC_NO:         101
```

```
124             LINE                     124

    ORDER_NO:     105
    LINE_NO:      003
    QTY:          2

    PART_NO:      021 REC_NO: 050
    NEXT LINE_NO REC_NO:      068
    NEXT ORDER_NO REC_NO:     072
```

```
125            LINE                       125

    ORDER_NO:   941
    LINE_NO:    001
    QTY:        1

    PART_NO:    450 REC_NO: 083
    NEXT LINE_NO REC_NO:    EOF
    NEXT ORDER_NO REC_NO:   EOF
```

```
126            PARTS                      126

    PART_NO:    546
    PART_NAME:  BRACKET
    EMPLO_NO:   354 REC_NO: 025

    SUPPL1_NO:  010 REC_NO: 049
    SUPPL2_NO:  065 REC_NO: 057
    NEXT PART_NO REC_NO:    106
    NEXT PART NAME REC_NO:  083
```

```
127            ORDER                      127

    ORDER_NO:   836
    CUST_NO:    564
    DATE:       29 AUG 88

    NEXT ORDER_NO REC_NO:   120
    THIS CUST_NO REC_NO:    024
    NEXT DATE REC_NO:       001
```

```
128            LINE                       128

    ORDER_NO:   922
    LINE_NO:    001
    QTY:        4

    PART_NO:    123 REC_NO: 019
    NEXT LINE_NO REC_NO:    EOF
    NEXT ORDER_NO REC_NO:   125
```

```
129            PARTS                         129

    PART_NO:     658
    PART_NAME:   PLUG
    EMPLO_NO:    443 REC_NO: 067

    SUPPL1_NO:   065 REC_NO: 057
    SUPPL2_NO:   021 REC_NO: 062
    NEXT PART_NO REC_NO:     132
    NEXT PART NAME REC_NO:   121
```

```
130            CUST                          130

    CUST_NO:     134
    CUST_NAME:   WELTY
    ADDRESS:     AMHERST

    NEXT CUST_NO REC_NO:     122
    NEXT CUST_NAME REC_NO:   027
```

```
131            LINE                          131

    ORDER_NO:    030
    LINE_NO:     001
    QTY:         4

    PART_NO:     240 REC_NO: 085
    NEXT LINE_NO REC_NO:     EOF
    NEXT ORDER_NO REC_NO:    086
```

```
132            PARTS                         132

    PART_NO:     718
    PART_NAME:   CAP
    EMPLO_NO:    064 REC_NO: 093

    SUPPL1_NO:   065 REC_NO: 057
    SUPPL2_NO:   043 REC_NO: 104
    NEXT PART_NO REC_NO:     121
    NEXT PART NAME REC_NO:   106
```

```
133              PARTS                       133

    PART_NO:      822
    PART_NAME:    BLADE
    EMPLO_NO:     127 REC_NO: 040

    SUPPL1_NO:    087 REC_NO: 065
    SUPPL2_NO:    065 REC_NO: 057
    NEXT PART_NO REC_NO:      052
    NEXT PART NAME REC_NO:    033
```

```
134              PARTS                       134

    PART_NO:      625
    PART_NAME:    GASKET
    EMPLO_NO:     443 REC_NO: 067

    SUPPL1_NO:    043 REC_NO: 104
    SUPPL2_NO:    021 REC_NO: 062
    NEXT PART_NO REC_NO:      129
    NEXT PART NAME REC_NO:    075
```

```
135              ORDER                       135

    ORDER_NO:     010
    CUST_NO:      376
    DATE:         15 DEC 88

    NEXT ORDER_NO REC_NO:     107
    THIS CUST_NO REC_NO:      048
    NEXT DATE REC_NO:         EOF
```

BIBLIOGRAPHY

Abrial, J.R., "Data Semantics", in: Klimbie, J.W. and Koffeman, K.L. (eds.) *Data Base Management*, North-Holland, 1974.

Bachman, C.W. and Daya, M., "The Role Concept in Data Models", *Proceedings of the International Conference on Very Large Data Bases*, Tokyo, October 1977.

Belkin, N. and Croft, B., "Retrieval Techniques", from University of Massachusetts working paper submitted to ARIST 87, 1987.

Benbasat, I. and Dexter, A.S., "An Experimental Evaluation of Graphical and Color-Enhanced Information Presentation," *Management Science*, November 1985, pp. 1348- 1364.

Benbasat, I., Dexter, A.S., and Masulis, P.S., "An Experimental Study of the Human/Computer Interface," *Communications of the ACM*, November 1981, pp. 752-762.

Benbasat, I., Dexter, A.S., and Todd, P., "An Experimental Program Investigating Color-Enhanced and Graphical Information Presentation: An investigation of the findings," *Communications of the ACM,* November 1986.

Benneworth, R.L., Bishop, C.D., Turnbull, C.J.M., Holman, W.D., and Monette, F.M., "The Implementation of GERM, an Entity-Relationship Data Base Management System", in: *Proceedings of the Seventh International Conference on Very Large Data Bases*, IEEE, 1981.

Bodart, F., "Opening Address", *Entity-Relationship Approach: Ten Years of Experience in Information Modeling*, Proceedings of the Fifth International Conference on Entity-Relationship Approach, Dijon, France, November 1986, North-Holland, 1987.

Broadbent, D.E. and Broadbent, M.H.P., "The allocation of descriptor terms by individuals in a simulated retrieval system," *Ergonomics*, 21, 1978, pp. 343-354.

Brodie, M.L., Mylopoulos, J., and Schmidt, J.W. (eds.), *On Conceptual Modeling*: Perspectives from Artificial Intelligence, Databases, and Programming Languages, Springer-Verlag, 1984. pp. 87-114.

Brosey, M. and Shneiderman, B., "Two Experimental Comparisons of Relational and Hierarchical Database Models", *International Journal of Man-Machine Studies*, 10, 1978, pp. 625-637.

Bubenko, J.A., "The Temporal Dimension in Information Modeling", *Architecture and Models in Data Base Management Systems*, Nijssen, G.M. (ed.) North-Holland, 1977.

Card, S., Moran, T.P., and Newell, A., *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, 1983.

Chamberlin, D.D., and Boyce, R.F., "SEQUEL: A Structured English Query Language", in *Proceedings of ACM SIGMOD,* 1974, pp. 249-264.

Chen, P.P.S., "The Entity-Relationship Model: Toward a Unified View of Data", *ACM TODS*, March 1976, pp. 9-36.

Chen, P.P.S., "Entity-Relationship diagrams and English sentence structures", (abstract only), in: Chen, P.P.S. (ed.) *Entity-Relationship Approach to System Analysis and Design*, North-Holland, 1980.

Chen, P.P.S., "A Preliminary Framework for Entity-Relationship Models", in: Chen, P.P.S., *The Entity-Relationship Approach to Information Modeling and Analysis*, E-R Institute, 1981.

Chen, P.P.S., "PC Based Productivity Tools for Data Modeling Professionals", Chen and Associates, Inc., 1988.

CODASYL, Data Base Task Group Report, ACM, 1971.

Codd, E.F., "A Relational Model of Data for Large shared Data Banks", *Communication of ACM*, June 1970, pp. 377-387.

Codd, E.F. and Date, C.J., "Interactive support for non-programmers: The relational and network approaches", *Proceedings of ACM SIGMOD 74*, ACM, 1974, pp. 11-41.

Codd, E.F., "Extending the Database Relational Model to Capture More Meaning", *ACM Transactions on Database Systems*, 1979, pp. 397-434.

Croft, W.B. and Stemple, D.W. "Supporting Office Document Architectures with Constrained Types", University of Massachusetts working paper, to have appeared in SIGMOD 87.

Date, C.J., *An Introduction to Database Systems, Volume II*, Addison-Wesley, 1985.

Data General / Database Management System (DG/DBMS) Reference Manual 093-000163-02, Data General Corporation, undated.

Dittrich, K.R., "Object-Oriented Database Systems: The Notion and the Issues - an extended abstract", *Proceedings of International Workshop on Object-Oriented Database Systems,* Dittrich, Klaus, Dayal, and Umeshwar, (eds.), IEEE Computer Society Press, 1986.

dos Santos, C.S., Neuhold, E.J. and Furtado, A.L., "A data type approach to the entity-relationship model", in: Chen, P.P.S. (ed.) *Entity-Relationship Approach to System Analysis and Design*, North-Holland, 1980.

Durding, B.M., Becker, C.A., and Gould, J.D., "Data organization", in *Human Factors*, 19, 1977, pp. 1-14.

Falkenberg, E. "Concepts for Modelling Information", in: Nijssen, G.M. (ed.) *Modelling in Data Base Management Systems*, North-Holland, 1976, pp. 95-110.

Flaaten, P.O., McCubbery, D.J., O'Riordan, P.D., and Burgess, K., *Foundations of Business Systems*, Dryden, 1989.

Furtado, A.L., Veloso, P.A.S., and de Castilho, J.M.V., "Verification and Testing of S-ER representations", in: Chen, P.P.S. (ed.) *Entity-Relationship Approach to Information Modeling and Analysis*, E-R Institute, 1981.

Gibbs, S. "An Object-Oriented Office Data Model." Ph.D. Thesis, University of Toronto, 1984.

Gold, J.A., *Principles of Psychological Research*, The Dorsey Press, 1984.

Greenblatt, D. and Waxman, J. "A study of three database query languages," in *Databases: Improving usability and responsiveness*, Shneiderman, B. (ed.) Academic Press, 1978.

Hammer, M. and McLoed, D., "Database Description with SDM: A Semantic Database Model", *ACM Transactions on Database Systems*, September 1981, pp. 351-386.

Harary, F. and Norman, R.Z., *Graph Theory*, University of Michigan Press, 1953.

Hoffer, J.A., "An Empirical Investigation into Individual Differences in Database Models", in: *Proceedings of the Third International Information Systems Conference*, Ann Arbor, 1982.

IBM, *Information Management System / Virtual Storage (IMS/VS)*, General Information Manual, GH20-1260-3, IBM Corp., 1975.

IBM, "Audit and Control in the DB2 Environment", IBM Corp., in: *Software News, November Supplement,* 1987.

INGRES Report, MBM 092C 001, Relational Technology, Inc., 1988.

Jones, S. and Mason, P.J., "Handling the Time Dimension in a Data Base" in: *Proceedings of the International Conference on Data Bases*, Scotland, 1980, Heyden and Son, 1980.

Junet, M., Falquet, G., and Leonard, M., "ECRINS/86: An Extended Entity-Relationship Data Base Management System and its Semantic Query Language", in: *Proceedings of the Twelfth International Conference on VLDB,* Kyoto, 1986, pp. 259-266.

Kent, W., *Data and Reality*, North-Holland, 1978.

Kerlinger, F.N., *Foundations of Behavioral Research*, Third Edition, 1986.

Kerschberg, L., Klug, A. and Tsichritzis, D., "A Taxonomy of Data Models", in: Lockeman, P.C., and Neuhold, E.J., (eds.) *Systems for Large Data Bases*, North-Holland, 1976, pp. 43-64.

Klopprogge, M.R., "TERM: An Approach to Include the Time Dimension in the Entity-Relationship Model", in: Chen, P.P.S., (ed.) *Entity-Relationship Approach to Information Modeling and Analysis*, E-R Institute, 1981, pp. 477-512.

Langefors, B., "Information Systems Theory", *Information Systems*, Vol. 2, 1977.

Lochovsky, F.H., "Data Base Management System User Performance", Ph.D. thesis, Dept. of Computer Science, University of Toronto, Toronto, 1978.

Lochovsky, F.H. and Tsichritzis, D.C., "An Interactive Query Language for External Data Bases", in: *Proceedings of the Eighth International Conference on Very Large Data Bases*, VLDB Endowment, 1982.

Lum, V.Y. et al., "1978 New Orleans Data Base Design Workshop Report", in: *Proceedings of the Fifth International Conference on Very Large Data Bases*, IEEE, 1979, pp. 328-339.

Lusk, E., Petrie, G., and Overbeek, R., "Item tracking entity-relationship models", in: Chen, P.P.S. (ed.), *Entity-Relationship Approach to Information Modeling and Analysis*, E-R Institute, 1981.

Mantei, M., *Disorientation Behavior in Person-Computer Interaction*, Ph.D. thesis, Annenberg School of Communications, University of Southern California, 1982.

Minsky, M.A., "A Framework of Representing Knowledge", in: Winston. P.H. (ed.), *The Psychology of Computer Vision*, McGraw-Hill, 1975.

Mylopoulos, J., Bernstein, P., and Wong, H.K.T., "A Preliminary report on TAXIS: A Language for Interactive Information System Design", Computer Corporation of America Report, CCA-77-10, December 1977.

Navathe, S. and Schkolnick, M., "View Representation in Logical Database Design", *ACM SIGMOD*, 1978.

Nolan, R.L. (ed.), *Managing the Data Resource Function*, West Publishing, 1974.

Quillian, M.R., "Semantic Memory", in: Minsky, M.A., *Semantic Information Processing*, MIT Press, 1968, pp. 227-268.

Ramaprasad, A., "Cognitive Process as a basis for MIS and DSS Design", *Management Science*, February 1987, pp. 139-148.

Ramsey, H.R. and Atwood, M.E., *Human Factors in Computer Systems: A Review of the Literature*, Science Application, Inc., September 1979.

Reiner, Brodie, Brown, Friedell, Kramlich, Lehman, and Rosenthal, "The Database Design and Evaluation Workbench (DDEW) Project at CCA", *IEEE Database Engineering* Vol. 7, No. 4, December 1984.

Reisner, P., "Use of psychological experimentation as an aid to development of a query language", *IEEE Transactions on Software Engineering 3*, 1977, pp. 218-229.

Reisner, P., "Human Factor Studies of Database Query Languages: A Survey and Assessment", *ACM Computing Surveys*, 13, 1981, pp. 13-31.

Reisner, P., Boyce, R.F., and Chamberlin, D.D., "Human factors evaluation of two data base query languages - SQUARE and SEQUEL", in: *Proceedings of American Interaction of Information Processing Societies National Computer Conference*, 1975, pp. 447-452.

Robinson, K.A., "An Entity / Event Data Modeling Method", *Computer Journal*, August, 1979.

Sakai, H., "A Method for Defining Information Structures and Transactions in Conceptual Schema Design", *Proceedings of the Seventh International Conference on Very Large Data Bases*, IEEE 1981, pp. 225-234.

Scheuermann, P., Schiffner, G. and Weber, H., "Abstraction Capabilities and Invariant Properties Modeling within the Entity-Relationship Approach", in: Chen, P.P.S. (ed.), *Entity-Relationship Approach to System Analysis and Design*, North-Holland, June 1980.

Schmid, H.A. and Swenson, J.R., "On the Semantics of the Relational Data Model", *Proceedings of the International Conference on Management of Data*, ACM SIGMOD, 1975, pp. 211-223.

Schueler, B.M., "Update Reconsidered", in *Architecture and Models in Data Base Management Systems*, North-Holland, 1977.

Senko, M.E., Altman, E.B., Astrahan, M.M. and Fehder, P.L., "Data Structures and Accessing in Data-Base Systems", *IBM Systems Journal*, 12,1 1973, p. 30-93.

Shipman, D.W., "The Functional Data Model and the Data Language DAPLEX", *ACM Transactions on Database Systems*, 6, 1981, pp. 140-173.

Shneiderman, B. "Improving the human factors aspect of database interactions", *ACM Transactions on Database Systems,* 3, 1978, pp. 417-439.

Shneiderman, B., *Designing the User Interface*: Strategies for Effective Human-Computer Interaction, Addison- Wesley, 1987.

Sibley, E.H., "On the Equivalence of Database Systems", in: *Proceedings of ACM SIGMOD 1974*: Debate on Data Models, ACM, May 1974, 43-76.

Sibley, E.H. and Kershberg, L., "Data Architecture and Data Model Considerations", in: *Proceedings of AFIPS National Computer Conference*, June 1977, p.85-96.

Smith, J.M., and Smith, D.C.P., "Database Abstractions: Aggregation and Generalization", *ACM Transactions on Database Systems*, June 1977, pp. 105-133.

Smith, J.M., and Smith, D.C.P., "Principles of Database Conceptual Design", *Proceedings of the NYU Symposium on Database Design*, New York University, 1978, pp. 35-49.

Sowa, J.F., *Conceptual Structures*, Addison-Wesley, 1984.

Stonebraker, M. and Rowe, L., "The Design of POSTGRES", *Proceedings of ACM SIGMOD 86*, ACM, 1986, pp. 340-355.

Su, S.Y.W. and Lo, D.H., "A Semantic Association Model for Conceptual Database Design", in: *Proceedings of the International Conference of The Entity-Relationship Approach to Systems Analysis and Design*, Los Angeles, E-R Institute, 1979.

Taylor, R.W. and Frank, R.L., "CODASYL Data Base Management Systems", *ACM Computing Surveys*, March 1986.

Thomas, J.C., and Gould, J.D.,"A Psychological Study of Query-by-Example", *Proceedings of the National Computer Conference*, AFIPS Press, Vol. 44, 1975, pp. 439-445.

Tsichritzis, D.C., and Lochovsky, F.H., *Data Models*, Prentice-Hall, 1982.

Tsichritzis, D.C. and Klug, A., (eds.) "The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems", *Information Systems*, 3, 1978, pp. 173-191.

Vassiliou, Y., (ed.) *Human Factors and Interactive Computer Systems*, Ablex, 1984.

Webster's New Collegiate Dictionary, G.& C. Merriman, 1975.

Welty, C. and Stemple, D.W., "Human Factors Comparison of a Procedural and a Non-procedural Query Language", *ACM Transactions on Database Systems*, Dec 1981.

Woelk, D., Kim, W. and Luther, W., "An Object-Oriented Approach to Multimedia Databases." in: *Proceedings of SIGMOD 86*, ACM, 1986, pp. 311-325.

Zloof, M.M., "Query by Example, The Invocation and Definition of Tables and Forms", *Proceedings of Conference of Very Large Data Bases,* ACM, Boston, MA, September 1975.

Zloof, M.M., "Security and Integrity within the Query-by-Example Data Base Management Language", IBM Research, working paper, RC6982, February 1978.